

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
18 October 2001 (18.10.2001)

PCT

(10) International Publication Number
WO 01/77174 A2

(51) International Patent Classification: C07K 14/705

(21) International Application Number: PCT/US01/11206

(22) International Filing Date: 6 April 2001 (06.04.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/195,595 6 April 2000 (06.04.2000) US
60/196,872 12 April 2000 (12.04.2000) US
60/199,020 20 April 2000 (20.04.2000) US
60/200,552 28 April 2000 (28.04.2000) US
60/202,348 5 May 2000 (05.05.2000) US
60/203,495 11 May 2000 (11.05.2000) US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:

US 60/195,595 (CIP)
Filed on 6 April 2000 (06.04.2000)
US 60/196,872 (CIP)
Filed on 12 April 2000 (12.04.2000)
US 60/199,020 (CIP)
Filed on 20 April 2000 (20.04.2000)
US 60/200,552 (CIP)
Filed on 28 April 2000 (28.04.2000)
US 60/202,348 (CIP)
Filed on 5 May 2000 (05.05.2000)
US 60/203,495 (CIP)
Filed on 11 May 2000 (11.05.2000)

(71) Applicant (for all designated States except US): INCYTE GENOMICS, INC. [US/US]; 3160 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): REDDY, Roopa [IN/US]; 1233 W. McKinley Avenue, #3, Sunnyvale, CA 94086 (US). THORNTON, Michael [US/US]; 9 Medway Road, Woodside, CA 94062-2612 (US). BOROWSKY, Mark, L. [US/US]; 122 Orchard Avenue, Redwood City, CA 94061 (US). TANG, Y., Tom [US/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). KHAN, Farrah, A. [IN/US]; 333 Escuela Avenue, #221, Mountain View, CA 94040 (US). TRIBOULEY, Catherine, M. [FR/US]; 1121 Tennessee Street, #5, San Francisco, CA 94107 (US). GANDHI, Ameena, R. [US/US]; 837 Roble Avenue, #1, Menlo Park, CA 94025 (US). YAO, Monique, G.

[US/US]; 111 Frederick Court, Mountain View, CA 94043 (US). SANJANWALA, Madhu, S. [US/US]; 210 Sylvia Court, Los Altos, CA 94024 (US). BAUGHN, Mariah, R. [US/US]; 14244 Santiago Road, San Leandro, CA 94577 (US). NGUYEN, Dannel, B. [US/US]; 1403 Ridgewood Drive, San Jose, CA 95118 (US). POLICKY, Jennifer, L. [US/US]; 1511 Jarvis Court, San Jose, CA 95118 (US). YUE, Henry [US/US]; 826 Lois Avenue, Sunnyvale, CA 94087 (US). SEILHAMER, Jeffrey, J. [US/US]; 12555 La Cresta, Los Altos Hills, CA 94022 (US). WALIA, Narinder, K. [US/US]; 890 Davis Street, #205, San Leandro, CA 94577 (US). LAL, Preeti [IN/US]; P.O. Box 5142, Santa Clara, CA 95056 (US). KEARNEY, Liam [IE/US]; 50 Woodside Avenue, San Francisco, CA 94127 (US). WALSH, Roderick, T. [IE/GB]; 8 Boundary Court, St. Lawrence Road, Canterbury, Kent CT1 3EZ (GB). LU, Dyung, Aina, M. [US/US]; 233 Coy Drive, San Jose, CA 95123 (US). LU, Yan [CN/US]; 3885 Corrina Way, Palo Alto, CA 94303 (US). GREENE, Barrie, D. [US/US]; 1332 10th Avenue, #104, San Francisco, CA 94122 (US). RAUMANN, Brigitte, E. [US/US]; 5801 South Dorchester Avenue, #3B, Chicago, IL 60637 (US). PATTERSON, Chandra [US/US]; 490 Sherwood Way, #1, Menlo Park, CA 94025 (US).

(74) Agents: HAMLET-COX, Diana et al.; Incyte Genomics, Inc., 3160 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, GR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MY, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, NI, SN, TD, TG).

Published:

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: TRANSPORTERS AND ION CHANNELS

(57) Abstract: The invention provides human transporters and ion channels (TRICH) and polynucleotides which identify and encode TRICH. The invention also provides expression vectors, host cells, antibodies, agonists, and antagonists. The invention also provides methods for diagnosing, treating, or preventing disorders associated with aberrant expression of TRICH.



WO 01/77174 A2

TRANSPORTERS AND ION CHANNELS

TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of transporters and ion channels and to the use of these sequences in the diagnosis, treatment, and prevention of transport, neurological, muscle, immunological, and cell proliferative disorders, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of transporters and ion channels.

BACKGROUND OF THE INVENTION

Eukaryotic cells are surrounded and subdivided into functionally distinct organelles by hydrophobic lipid bilayer membranes which are highly impermeable to most polar molecules. Cells and organelles require transport proteins to import and export essential nutrients and metal ions including K^+ , NH_4^+ , P_i , SO_4^{2-} , sugars, and vitamins, as well as various metabolic waste products. Transport proteins also play roles in antibiotic resistance, toxin secretion, ion balance, synaptic neurotransmission, kidney function, intestinal absorption, tumor growth, and other diverse cell functions (Griffith, J. and C. Sansom (1998) The Transporter Facts Book, Academic Press, San Diego CA, pp. 3-29). Transport can occur by a passive concentration-dependent mechanism, or can be linked to an energy source such as ATP hydrolysis or an ion gradient. Proteins that function in transport include carrier proteins, which bind to a specific solute and undergo a conformational change that translocates the bound solute across the membrane, and channel proteins, which form hydrophilic pores that allow specific solutes to diffuse through the membrane down an electrochemical solute gradient.

Carrier proteins which transport a single solute from one side of the membrane to the other are called uniporters. In contrast, coupled transporters link the transfer of one solute with simultaneous or sequential transfer of a second solute, either in the same direction (symport) or in the opposite direction (antiport). For example, intestinal and kidney epithelium contains a variety of symporter systems driven by the sodium gradient that exists across the plasma membrane. Sodium moves into the cell down its electrochemical gradient and brings the solute into the cell with it. The sodium gradient that provides the driving force for solute uptake is maintained by the ubiquitous Na^+/K^+ ATPase system. Sodium-coupled transporters include the mammalian glucose transporter (SGLT1), iodide transporter (NIS), and multivitamin transporter (SMVT). All three transporters have twelve putative transmembrane segments, extracellular glycosylation sites, and cytoplasmically-oriented N- and C-termini. NIS plays a crucial role in the evaluation, diagnosis, and treatment of various thyroid pathologies because it is the molecular basis for radioiodide thyroid-imaging

techniques and for specific targeting of radioisotopes to the thyroid gland (Levy, O. et al. (1997) Proc. Natl. Acad. Sci. USA 94:5568-5573). SMVT is expressed in the intestinal mucosa, kidney, and placenta, and is implicated in the transport of the water-soluble vitamins, e.g., biotin and pantothenate (Prasad, P.D. et al. (1998) J. Biol. Chem. 273:7501-7506).

- 5 One of the largest families of transporters is the major facilitator superfamily (MFS), also called the uniporter-symporter-antiporter family. MFS transporters are single polypeptide carriers that transport small solutes in response to ion gradients. Members of the MFS are found in all classes of living organisms, and include transporters for sugars, oligosaccharides, phosphates, nitrates, nucleosides, monocarboxylates, and drugs. MFS transporters found in eukaryotes all have a structure
- 10 comprising 12 transmembrane segments (Pao, S.S. et al. (1998) Microbiol. Molec. Biol. Rev. 62:1-34). The largest family of MFS transporters is the sugar transporter family, which includes the seven glucose transporters (GLUT1-GLUT7) found in humans that are required for the transport of glucose and other hexose sugars. These glucose transport proteins have unique tissue distributions and physiological functions. GLUT1 provides many cell types with their basal glucose requirements and
- 15 transports glucose across epithelial and endothelial barrier tissues; GLUT2 facilitates glucose uptake or efflux from the liver; GLUT3 regulates glucose supply to neurons; GLUT4 is responsible for insulin-regulated glucose disposal; and GLUT5 regulates fructose uptake into skeletal muscle. Defects in glucose transporters are involved in a recently identified neurological syndrome causing infantile seizures and developmental delay, as well as glycogen storage disease, Fanconi-Bickel
- 20 syndrome, and non-insulin-dependent diabetes mellitus (Mueckler, M. (1994) Eur. J. Biochem. 219:713-725; Longo, N. and L.J. Elsas (1998) Adv. Pediatr. 45:293-313).

- Monocarboxylate anion transporters are proton-coupled symporters with a broad substrate specificity that includes L-lactate, pyruvate, and the ketone bodies acetate, acetoacetate, and beta-hydroxybutyrate. At least seven isoforms have been identified to date. The isoforms are
- 25 predicted to have twelve transmembrane (TM) helical domains with a large intracellular loop between TM6 and TM7, and play a critical role in maintaining intracellular pH by removing the protons that are produced stoichiometrically with lactate during glycolysis. The best characterized H⁺-monocarboxylate transporter is that of the erythrocyte membrane, which transports L-lactate and a wide range of other aliphatic monocarboxylates. Other cells possess H⁺-linked monocarboxylate
- 30 transporters with differing substrate and inhibitor selectivities. In particular, cardiac muscle and tumor cells have transporters that differ in their K_m values for certain substrates, including stereoselectivity for L- over D-lactate, and in their sensitivity to inhibitors. There are Na⁺-monocarboxylate cotransporters on the luminal surface of intestinal and kidney epithelia, which allow the uptake of lactate, pyruvate, and ketone bodies in these tissues. In addition, there are
- 35 specific and selective transporters for organic cations and organic anions in organs including the

kidney, intestine and liver. Organic anion transporters are selective for hydrophobic, charged molecules with electron-attracting side groups. Organic cation transporters, such as the ammonium transporter, mediate the secretion of a variety of drugs and endogenous metabolites, and contribute to the maintenance of intercellular pH (Poole, R.C. and A.P. Halestrap (1993) *Am. J. Physiol.*

- 5 264:C761-C782; Price, N.T. et al. (1998) *Biochem. J.* 329:321-328; and Martinelle, K. and I. Haggstrom (1993) *J. Biotechnol.* 30:339-350).

ATP-binding cassette (ABC) transporters are members of a superfamily of membrane proteins that transport substances ranging from small molecules such as ions, sugars, amino acids, peptides, and phospholipids, to lipopeptides, large proteins, and complex hydrophobic drugs. ABC
10 transporters consist of four modules: two nucleotide-binding domains (NBD), which hydrolyze ATP to supply the energy required for transport, and two membrane-spanning domains (MSD), each containing six putative transmembrane segments. These four modules may be encoded by a single gene, as is the case for the cystic fibrosis transmembrane regulator (CFTR), or by separate genes. When encoded by separate genes, each gene product contains a single NBD and MSD. These "half-
15 molecules" form homo- and heterodimers, such as Tap1 and Tap2, the endoplasmic reticulum-based major histocompatibility (MHC) peptide transport system. Several genetic diseases are attributed to defects in ABC transporters, such as the following diseases and their corresponding proteins: cystic fibrosis (CFTR, an ion channel), adrenoleukodystrophy (adrenoleukodystrophy protein, ALDP), Zellweger syndrome (peroxisomal membrane protein-70, PMP70), and hyperinsulinemic
20 hypoglycemia (sulfonylurea receptor, SUR). Overexpression of the multidrug resistance (MDR) protein, another ABC transporter, in human cancer cells makes the cells resistant to a variety of cytotoxic drugs used in chemotherapy (Taglicht, D. and S. Michaelis (1998) *Meth. Enzymol.* 292:130-162).

A number of metal ions such as iron, zinc, copper, cobalt, manganese, molybdenum,
25 selenium, nickel, and chromium are important as cofactors for a number of enzymes. For example, copper is involved in hemoglobin synthesis, connective tissue metabolism, and bone development, by acting as a cofactor in oxidoreductases such as superoxide dismutase, ferroxidase (ceruloplasmin), and lysyl oxidase. Copper and other metal ions must be provided in the diet, and are absorbed by transporters in the gastrointestinal tract. Plasma proteins transport the metal ions to the liver and
30 other target organs, where specific transporters move the ions into cells and cellular organelles as needed. Imbalances in metal ion metabolism have been associated with a number of disease states (Danks, D.M. (1986) *J. Med. Genet.* 23:99-106).

Transport of fatty acids across the plasma membrane can occur by diffusion, a high capacity, low affinity process. However, under normal physiological conditions a significant fraction of fatty
35 acid transport appears to occur via a high affinity, low capacity protein-mediated transport process.

Fatty acid transport protein (FATP), an integral membrane protein with four transmembrane segments, is expressed in tissues exhibiting high levels of plasma membrane fatty acid flux, such as muscle, heart, and adipose. Expression of FATP is upregulated in 3T3-L1 cells during adipose conversion, and expression in COS7 fibroblasts elevates uptake of long-chain fatty acids (Hui, T.Y. et

al. (1998) *J. Biol. Chem.* 273:27420-27429).

Mitochondrial carrier proteins are transmembrane-spanning proteins which transport ions and charged metabolites between the cytosol and the mitochondrial matrix. Examples include the ADP, ATP carrier protein; the 2-oxoglutarate/malate carrier; the phosphate carrier protein; the pyruvate carrier; the dicarboxylate carrier which transports malate, succinate, fumarate, and phosphate; the tricarboxylate carrier which transports citrate and malate; and the Grave's disease carrier protein, a protein recognized by IgG in patients with active Grave's disease, an autoimmune disorder resulting in hyperthyroidism. Proteins in this family consist of three tandem repeats of an approximately 100 amino acid domain, each of which contains two transmembrane regions (Stryer, L. (1995)

Biochemistry, W.H. Freeman and Company, New York NY, p. 551; PROSITE PDOC00189

15 Mitochondrial energy transfer proteins signature; Online Mendelian Inheritance in Man (OMIM) *275000 Graves Disease).

This class of transporters also includes the mitochondrial uncoupling proteins, which create proton leaks across the inner mitochondrial membrane, thus uncoupling oxidative phosphorylation from ATP synthesis. The result is energy dissipation in the form of heat. Mitochondrial uncoupling proteins have been implicated as modulators of thermoregulation and metabolic rate, and have been proposed as potential targets for drugs against metabolic diseases such as obesity (Ricquier, D. et al. (1999) *J. Int. Med.* 245:637-642).

Ion Channels

The electrical potential of a cell is generated and maintained by controlling the movement of ions across the plasma membrane. The movement of ions requires ion channels, which form ion-selective pores within the membrane. There are two basic types of ion channels, ion transporters and gated ion channels. Ion transporters utilize the energy obtained from ATP hydrolysis to actively transport an ion against the ion's concentration gradient. Gated ion channels allow passive flow of an ion down the ion's electrochemical gradient under restricted conditions. Together, these types of ion channels generate, maintain, and utilize an electrochemical gradient that is used in 1) electrical impulse conduction down the axon of a nerve cell, 2) transport of molecules into cells against concentration gradients, 3) initiation of muscle contraction, and 4) endocrine cell secretion.

Ion Transporters

Ion transporters generate and maintain the resting electrical potential of a cell. Utilizing the energy derived from ATP hydrolysis, they transport ions against the ion's concentration gradient.

These transmembrane ATPases are divided into three families. The phosphorylated (P) class ion transporters, including Na⁺-K⁺ ATPase, Ca²⁺-ATPase, and H⁺-ATPase, are activated by a phosphorylation event. P-class ion transporters are responsible for maintaining resting potential distributions such that cytosolic concentrations of Na⁺ and Ca²⁺ are low and cytosolic concentration of K⁺ is high. The vacuolar (V) class of ion transporters includes H⁺ pumps on intracellular organelles, such as lysosomes and Golgi. V-class ion transporters are responsible for generating the low pH within the lumen of these organelles that is required for function. The coupling factor (F) class consists of H⁺ pumps in the mitochondria. F-class ion transporters utilize a proton gradient to generate ATP from ADP and inorganic phosphate (P_i).

The P-ATPases are hexamers of a 100 kD subunit with ten transmembrane domains and several large cytoplasmic regions that may play a role in ion binding (Scarborough, G.A. (1999) *Curr. Opin. Cell Biol.* 11:517-522). The V-ATPases are composed of two functional domains: the V₁ domain, a peripheral complex responsible for ATP hydrolysis; and the V₀ domain, an integral complex responsible for proton translocation across the membrane. The F-ATPases are structurally and evolutionarily related to the V-ATPases. The F-ATPase F₀ domain contains 12 copies of the c subunit, a highly hydrophobic protein composed of two transmembrane domains and containing a single buried carboxyl group in TM2 that is essential for proton transport. The V-ATPase V₀ domain contains three types of homologous c subunits with four or five transmembrane domains and the essential carboxyl group in TM4 or TM3. Both types of complex also contain a single a subunit that may be involved in regulating the pH dependence of activity (Forgacs, M. (1999) *J. Biol. Chem.* 274:12951-12954).

The resting potential of the cell is utilized in many processes involving carrier proteins and gated ion channels. Carrier proteins utilize the resting potential to transport molecules into and out of the cell. Amino acid and glucose transport into many cells is linked to sodium ion co-transport (symport) so that the movement of Na⁺ down an electrochemical gradient drives transport of the other molecule up a concentration gradient. Similarly, cardiac muscle links transfer of Ca²⁺ out of the cell with transport of Na⁺ into the cell (antiport).

Gated Ion Channels

Gated ion channels control ion flow by regulating the opening and closing of pores. The ability to control ion flux through various gating mechanisms allows ion channels to mediate such diverse signaling and homeostatic functions as neuronal and endocrine signaling, muscle contraction, fertilization, and regulation of ion and pH balance. Gated ion channels are categorized according to the manner of regulating the gating function. Mechanically-gated channels open their pores in response to mechanical stress; voltage-gated channels (e.g., Na⁺, K⁺, Ca²⁺, and Cl⁻ channels) open their pores in response to changes in membrane potential; and ligand-gated channels (e.g.,

acetylcholine-, serotonin-, and glutamate-gated cation channels, and GABA- and glycine-gated chloride channels) open their pores in the presence of a specific ion, nucleotide, or neurotransmitter. The gating properties of a particular ion channel (i.e., its threshold for and duration of opening and closing) are sometimes modulated by association with auxiliary channel proteins and/or post

5 translational modifications, such as phosphorylation.

Mechanically-gated or mechanosensitive ion channels act as transducers for the senses of touch, hearing, and balance, and also play important roles in cell volume regulation, smooth muscle contraction, and cardiac rhythm generation. A stretch-inactivated channel (SIC) was recently cloned from rat kidney. The SIC channel belongs to a group of channels which are activated by pressure or

10 stress on the cell membrane and conduct both Ca^{2+} and Na^{+} (Suzuki, M. et al. (1999) J. Biol. Chem. 274:6330-6335).

The pore-forming subunits of the voltage-gated cation channels form a superfamily of ion channel proteins. The characteristic domain of these channel proteins comprises six transmembrane domains (S1-S6), a pore-forming region (P) located between S5 and S6, and intracellular amino and

15 carboxy termini. In the Na^{+} and Ca^{2+} subfamilies, this domain is repeated four times, while in the K^{+} channel subfamily, each channel is formed from a tetramer of either identical or dissimilar subunits. The P region contains information specifying the ion selectivity for the channel. In the case of K^{+} channels, a GYG tripeptide is involved in this selectivity (Ishii, T.M. et al. (1997) Proc. Natl. Acad. Sci. USA 94:11651-11656).

Voltage-gated Na^{+} and K^{+} channels are necessary for the function of electrically excitable cells, such as nerve and muscle cells. Action potentials, which lead to neurotransmitter release and muscle contraction, arise from large, transient changes in the permeability of the membrane to Na^{+} and K^{+} ions. Depolarization of the membrane beyond the threshold level opens voltage-gated Na^{+} channels. Sodium ions flow into the cell, further depolarizing the membrane and opening more

20 voltage-gated Na^{+} channels, which propagates the depolarization down the length of the cell. Depolarization also opens voltage-gated potassium channels. Consequently, potassium ions flow outward, which leads to repolarization of the membrane. Voltage-gated channels utilize charged residues in the fourth transmembrane segment (S4) to sense voltage change. The open state lasts only about 1 millisecond, at which time the channel spontaneously converts into an inactive state that

30 cannot be opened irrespective of the membrane potential. Inactivation is mediated by the channel's N-terminus, which acts as a plug that closes the pore. The transition from an inactive to a closed state requires a return to resting potential.

Voltage-gated Na^{+} channels are heterotrimeric complexes composed of a 260 kDa pore-forming α subunit that associates with two smaller auxiliary subunits, $\beta 1$ and $\beta 2$. The $\beta 2$ subunit is a

35 integral membrane glycoprotein that contains an extracellular Ig domain, and its association with α

and $\beta 1$ subunits correlates with increased functional expression of the channel, a change in its gating properties, as well as an increase in whole cell capacitance due to an increase in membrane surface area (Isom, L.L. et al. (1995) Cell 83:433-442).

Non voltage-gated Na^+ channels include the members of the amiloride-sensitive Na^+ channel/degenerin (NaC/DEG) family. Channel subunits of this family are thought to consist of two transmembrane domains flanking a long extracellular loop, with the amino and carboxyl termini located within the cell. The NaC/DEG family includes the epithelial Na^+ channel (ENaC) involved in Na^+ reabsorption in epithelia including the airway, distal colon, cortical collecting duct of the kidney, and exocrine duct glands. Mutations in ENaC result in pseudohypoaldosteronism type 1 and Liddle's syndrome (pseudohyperaldosteronism). The NaC/DEG family also includes the recently characterized H^+ -gated cation channels or acid-sensing ion channels (ASIC). ASIC subunits are expressed in the brain and form heteromultimeric Na^+ -permeable channels. These channels require acid pH fluctuations for activation. ASIC subunits show homology to the degenerins, a family of mechanically-gated channels originally isolated from *C. elegans*. Mutations in the degenerins cause neurodegeneration. ASIC subunits may also have a role in neuronal function, or in pain perception, since tissue acidosis causes pain (Waldmann, R. and M. Lazdunski (1998) Curr. Opin. Neurobiol. 8:418-424; Eglén, R.M. et al. (1999) Trends Pharmacol. Sci. 20:337-342).

K^+ channels are located in all cell types, and may be regulated by voltage, ATP concentration, or second messengers such as Ca^{2+} and cAMP. In non-excitabile tissue, K^+ channels are involved in protein synthesis, control of endocrine secretions, and the maintenance of osmotic equilibrium across membranes. In neurons and other excitable cells, in addition to regulating action potentials and repolarizing membranes, K^+ channels are responsible for setting resting membrane potential. The cytosol contains non-diffusible anions and, to balance this net negative charge, the cell contains a Na^+ - K^+ pump and ion channels that provide the redistribution of Na^+ , K^+ , and Cl^- . The pump actively transports Na^+ out of the cell and K^+ into the cell in a 3:2 ratio. Ion channels in the plasma membrane allow K^+ and Cl^- to flow by passive diffusion. Because of the high negative charge within the cytosol, Cl^- flows out of the cell. The flow of K^+ is balanced by an electromotive force pulling K^+ into the cell, and a K^+ concentration gradient pushing K^+ out of the cell. Thus, the resting membrane potential is primarily regulated by K^+ flow (Salkoff, L. and T. Jegla (1995) Neuron 15:489-492).

Potassium channel subunits of the Shaker-like superfamily all have the characteristic six transmembrane/1 pore domain structure. Four subunits combine as homo- or heterotetramers to form functional K channels. These pore-forming subunits also associate with various cytoplasmic β subunits that alter channel inactivation kinetics. The Shaker-like channel family includes the voltage-gated K^+ channels as well as the delayed rectifier type channels such as the human ether-a-go-go

related gene (HERG) associated with long QT, a cardiac dysrhythmia syndrome (Curran, M.E. (1998) *Curr. Opin. Biotechnol.* 9:565-572; Kaczorowski, G.J. and M.L. Garcia (1999) *Curr. Opin. Chem. Biol.* 3:448-458).

A second superfamily of K⁺ channels is composed of the inward rectifying channels (Kir).

- 5 Kir channels have the property of preferentially conducting K⁺ currents in the inward direction. These proteins consist of a single potassium selective pore domain and two transmembrane domains, which correspond to the fifth and sixth transmembrane domains of voltage-gated K⁺ channels. Kir subunits also associate as tetramers. The Kir family includes ROMK1, mutations in which lead to Bartter syndrome, a renal tubular disorder. Kir channels are also involved in regulation of cardiac
- 10 pacemaker activity, seizures and epilepsy, and insulin regulation (Doupnik, C.A. et al. (1995) *Curr. Opin. Neurobiol.* 5:268-277; Curran, *supra*).

- The recently recognized TWIK K⁺ channel family includes the mammalian TWIK-1, TREK-1 and TASK proteins. Members of this family possess an overall structure with four transmembrane domains and two P domains. These proteins are probably involved in controlling the resting potential
- 15 in a large set of cell types (Duprat, F. et al. (1997) *EMBO J* 16:5464-5471).

- The voltage-gated Ca²⁺ channels have been classified into several subtypes based upon their electrophysiological and pharmacological characteristics. L-type Ca²⁺ channels are predominantly expressed in heart and skeletal muscle where they play an essential role in excitation-contraction coupling. T-type channels are important for cardiac pacemaker activity, while N-type and P/Q-type
- 20 channels are involved in the control of neurotransmitter release in the central and peripheral nervous system. The L-type and N-type voltage-gated Ca²⁺ channels have been purified and, though their functions differ dramatically, they have similar subunit compositions. The channels are composed of three subunits. The α_1 subunit forms the membrane pore and voltage sensor, while the α_2 , δ and β subunits modulate the voltage-dependence, gating properties, and the current amplitude of the
 - 25 channel. These subunits are encoded by at least six α_1 , one α_2 , and four β genes. A fourth subunit, γ , has been identified in skeletal muscle (Walker, D. et al. (1998) *J. Biol. Chem.* 273:2361-2367; McCleskey, E.W. (1994) *Curr. Opin. Neurobiol.* 4:304-312).

- The transient receptor family (Trp) of calcium ion channels are thought to mediate capacitative calcium entry (CCE). CCE is the Ca²⁺ influx into cells to resupply Ca²⁺ stores depleted
- 30 by the action of inositol triphosphate (IP3) and other agents in response to numerous hormones and growth factors. Trp and Trp-like were first cloned from *Drosophila* and have similarity to voltage gated Ca²⁺ channels in the S3 through S6 regions. This suggests that Trp and/or related proteins may form mammalian CCC entry channels (Zhu, X. et al. (1996) *Cell* 85:661-671; Boulay, G. et al. (1997) *J. Biol. Chem.* 272:29672-29680). Melastatin is a gene isolated in both the mouse and human, and
 - 35 whose expression in melanoma cells is inversely correlated with melanoma aggressiveness in vivo.

The human cDNA transcript corresponds to a 1533-amino acid protein having homology to members of the Trp family. It has been proposed that the combined use of malastatin mRNA expression status and tumor thickness might allow for the determination of subgroups of patients at both low and high risk for developing metastatic disease (Duncan, L.M. et al (2001) *J. Clin. Oncol.* 19:568-576).

- 5 Chloride channels are necessary in endocrine secretion and in regulation of cytosolic and organelle pH. In secretory epithelial cells, Cl^- enters the cell across a basolateral membrane through an Na^+ , K^+/Cl^- cotransporter, accumulating in the cell above its electrochemical equilibrium concentration. Secretion of Cl^- from the apical surface, in response to hormonal stimulation, leads to flow of Na^+ and water into the secretory lumen. The cystic fibrosis transmembrane conductance
- 10 regulator (CFTR) is a chloride channel encoded by the gene for cystic fibrosis, a common fatal genetic disorder in humans. CFTR is a member of the ABC transporter family, and is composed of two domains each consisting of six transmembrane domains followed by a nucleotide-binding site. Loss of CFTR function decreases transepithelial water secretion and, as a result, the layers of mucus that coat the respiratory tree, pancreatic ducts, and intestine are dehydrated and difficult to clear. The
- 15 resulting blockage of these sites leads to pancreatic insufficiency, "meconium ileus", and devastating "chronic obstructive pulmonary disease" (Al-Awqati, Q. et al. (1992) *J. Exp. Biol.* 172:245-266).

- The voltage-gated chloride channels (CLC) are characterized by 10-12 transmembrane domains, as well as two small globular domains known as CBS domains. The CLC subunits probably function as homotetramers. CLC proteins are involved in regulation of cell volume,
- 20 membrane potential stabilization, signal transduction, and transepithelial transport. Mutations in CLC-1, expressed predominantly in skeletal muscle, are responsible for autosomal recessive generalized myotonia and autosomal dominant myotonia congenita, while mutations in the kidney channel CLC-5 lead to kidney stones (Jentsch, T.J. (1996) *Curr. Opin. Neurobiol.* 6:303-310).

- Ligand-gated channels open their pores when an extracellular or intracellular mediator binds
- 25 to the channel. Neurotransmitter-gated channels are channels that open when a neurotransmitter binds to their extracellular domain. These channels exist in the postsynaptic membrane of nerve or muscle cells. There are two types of neurotransmitter-gated channels. Sodium channels open in response to excitatory neurotransmitters, such as acetylcholine, glutamate, and serotonin. This opening causes an influx of Na^+ and produces the initial localized depolarization that activates the
- 30 voltage-gated channels and starts the action potential. Chloride channels open in response to inhibitory neurotransmitters, such as γ -aminobutyric acid (GABA) and glycine, leading to hyperpolarization of the membrane and the subsequent generation of an action potential. Neurotransmitter-gated ion channels have four transmembrane domains and probably function as pentamers (Jentsch, *supra*). Amino acids in the second transmembrane domain appear to be important
- 35 in determining channel permeation and selectivity (Sather, W.A. et al. (1994) *Curr. Opin. Neurobiol.*

4:313-323).

Ligand-gated channels can be regulated by intracellular second messengers. For example, calcium-activated K^+ channels are gated by internal calcium ions. In nerve cells, an influx of calcium during depolarization opens K^+ channels to modulate the magnitude of the action potential (Ishi et al., supra). The large conductance (BK) channel has been purified from brain and its subunit composition determined. The α subunit of the BK channel has seven rather than six transmembrane domains in contrast to voltage-gated K^+ channels. The extra transmembrane domain is located at the subunit N-terminus. A 28-amino-acid stretch in the C-terminal region of the subunit (the "calcium bowl" region) contains many negatively charged residues and is thought to be the region responsible for calcium binding. The β subunit consists of two transmembrane domains connected by a glycosylated extracellular loop, with intracellular N- and C-termini (Kaczorowski, supra; Vergara, C. et al. (1998) *Curr. Opin. Neurobiol.* 8:321-329).

Cyclic nucleotide-gated (CNG) channels are gated by cytosolic cyclic nucleotides. The best examples of these are the cAMP-gated Na^+ channels involved in olfaction and the cGMP-gated cation channels involved in vision. Both systems involve ligand-mediated activation of a G-protein coupled receptor which then alters the level of cyclic nucleotide within the cell. CNG channels also represent a major pathway for Ca^{2+} entry into neurons, and play roles in neuronal development and plasticity. CNG channels are tetramers containing at least two types of subunits, an α subunit which can form functional homomeric channels, and a β subunit, which modulates the channel properties. All CNG subunits have six transmembrane domains and a pore forming region between the fifth and sixth transmembrane domains, similar to voltage-gated K^+ channels. A large C-terminal domain contains a cyclic nucleotide binding domain, while the N-terminal domain confers variation among channel subtypes (Zufall, F. et al. (1997) *Curr. Opin. Neurobiol.* 7:404-412).

The activity of other types of ion channel proteins may also be modulated by a variety of intracellular signalling proteins. Many channels have sites for phosphorylation by one or more protein kinases including protein kinase A, protein kinase C, tyrosine kinase, and casein kinase II, all of which regulate ion channel activity in cells. Kir channels are activated by the binding of the $G\beta\gamma$ subunits of heterotrimeric G-proteins (Reimann, F. and F.M. Ashcroft (1999) *Curr. Opin. Cell. Biol.* 11:503-508). Other proteins are involved in the localization of ion channels to specific sites in the cell membrane. Such proteins include the PDZ domain proteins known as MAGUKs (membrane-associated guanylate kinases) which regulate the clustering of ion channels at neuronal synapses (Craven, S.E. and D.S. Bredt (1998) *Cell* 93:495-498).

Disease Correlation

The etiology of numerous human diseases and disorders can be attributed to defects in the

transport of molecules across membranes. Defects in the trafficking of membrane-bound transporters and ion channels are associated with several disorders, e.g., cystic fibrosis, glucose-galactose malabsorption syndrome, hypercholesterolemia, von Gierke disease, and certain forms of diabetes mellitus. Single-gene defect diseases resulting in an inability to transport small molecules across
 5 membranes include, e.g., cystinuria, iminoglycinuria, Hartup disease, and Fanconi disease (van't Hoff, W.G. (1996) *Exp. Nephrol.* 4:253-262; Talente, G.M. et al. (1994) *Ann. Intern. Med.* 120:218-226; and Chillon, M. et al. (1995) *New Engl. J. Med.* 332:1475-1480).

Human diseases caused by mutations in ion channel genes include disorders of skeletal muscle, cardiac muscle, and the central nervous system. Mutations in the pore-forming subunits of
 10 sodium and chloride channels cause myotonia, a muscle disorder in which relaxation after voluntary contraction is delayed. Sodium channel myotonias have been treated with channel blockers. Mutations in muscle sodium and calcium channels cause forms of periodic paralysis, while mutations in the sarcoplasmic calcium release channel, T-tubule calcium channel, and muscle sodium channel cause malignant hyperthermia. Cardiac arrhythmia disorders such as the long QT syndromes and
 15 idiopathic ventricular fibrillation are caused by mutations in potassium and sodium channels (Cooper, E.C. and L.Y. Jan (1998) *Proc. Natl. Acad. Sci. USA* 96:4759-4766). All four known human idiopathic epilepsy genes code for ion channel proteins (Berkovic, S.F. and J.E. Scheffer (1999) *Curr. Opin. Neurology* 12:177-182). Other neurological disorders such as ataxias, hemiplegic migraine and hereditary deafness can also result from mutations in ion channel genes (Jen, J. (1999) *Curr. Opin.*
 20 *Neurobiol.* 9:274-280; Cooper, *supra*).

Ion channels have been the target for many drug therapies. Neurotransmitter-gated channels have been targeted in therapies for treatment of insomnia, anxiety, depression, and schizophrenia. Voltage-gated channels have been targeted in therapies for arrhythmia, ischemic stroke, head trauma, and neurodegenerative disease (Taylor, C.P. and L.S. Narasimhan (1997) *Adv. Pharmacol.* 39:47-98).
 25 Various classes of ion channels also play an important role in the perception of pain, and thus are potential targets for new analgesics. These include the vanilloid-gated ion channels, which are activated by the vanilloid capsaicin, as well as by noxious heat. Local anesthetics such as lidocaine and mexiletine which blockade voltage-gated Na⁺ channels have been useful in the treatment of neuropathic pain (Eglen, *supra*).

30 Ion channels in the immune system have recently been suggested as targets for immunomodulation. T-cell activation depends upon calcium signaling, and a diverse set of T-cell specific ion channels has been characterized that affect this signaling process. Channel blocking agents can inhibit secretion of lymphokines, cell proliferation, and killing of target cells. A peptide antagonist of the T-cell potassium channel Kv1.3 was found to suppress delayed-type hypersensitivity
 35 and allogenic responses in pigs, validating the idea of channel blockers as safe and efficacious

immunosuppressants (Cahalan, M.D. and K.G. Chandy (1997) Curr. Opin. Biotechnol. 8:749-756).

The discovery of new transporters and ion channels and the polynucleotides encoding them satisfies a need in the art by providing new compositions which are useful in the diagnosis, prevention, and treatment of transport, neurological, muscle, immunological, and cell proliferative disorders, and in the assessment of the effects of exogenous compounds on the expression of nucleic acid and amino acid sequences of transporters and ion channels.

SUMMARY OF THE INVENTION

The invention features purified polypeptides, transporters and ion channels, referred to collectively as "TRICH" and individually as "TRICH-1," "TRICH-2," "TRICH-3," "TRICH-4," "TRICH-5," "TRICH-6," "TRICH-7," "TRICH-8," "TRICH-9," "TRICH-10," "TRICH-11," "TRICH-12," "TRICH-13," "TRICH-14," and "TRICH-15." In one aspect, the invention provides an isolated polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15. In one alternative, the invention provides an isolated polypeptide comprising the amino acid sequence of SEQ ID NO:1-15.

The invention further provides an isolated polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15. In one alternative, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-15. In another alternative, the polynucleotide is selected from the group consisting of SEQ ID NO:16-30.

Additionally, the invention provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, c) a

biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15. In one alternative, the invention provides a cell transformed with the recombinant polynucleotide. In another alternative, the invention provides a transgenic organism comprising the recombinant polynucleotide.

The invention also provides a method for producing a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

Additionally, the invention provides an isolated antibody which specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15.

The invention further provides an isolated polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:16-30, b) a naturally occurring polynucleotide comprising a polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:16-30, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). In one alternative, the polynucleotide comprises at least 60 contiguous nucleotides.

Additionally, the invention provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:16-30, b) a naturally occurring polynucleotide comprising a polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of

SEQ ID NO:16-30, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and
5 which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex, and optionally, if present, the amount thereof. In one alternative, the probe comprises at least 60 contiguous nucleotides.

10 The invention further provides a method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:16-30, b) a naturally occurring polynucleotide comprising a polynucleotide sequence at
15 least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:16-30, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

20 The invention further provides a composition comprising an effective amount of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, c) a biologically active fragment of a polypeptide having an
25 amino acid sequence selected from the group consisting of SEQ ID NO:1-15, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, and a pharmaceutically acceptable excipient. In one embodiment, the composition comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-15. The invention additionally provides a method of treating a disease or condition associated with decreased
30 expression of functional TRICH, comprising administering to a patient in need of such treatment the composition.

The invention also provides a method for screening a compound for effectiveness as an agonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, b) a naturally occurring
35 polypeptide comprising an amino acid sequence at least 90% identical to an amino acid sequence

selected from the group consisting of SEQ ID NO:1-15, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. In one alternative, the invention provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with decreased expression of functional TRICH, comprising administering to a patient in need of such treatment the composition.

Additionally, the invention provides a method for screening a compound for effectiveness as an antagonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. In one alternative, the invention provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. In another alternative, the invention provides a method of treating a disease or condition associated with overexpression of functional TRICH, comprising administering to a patient in need of such treatment the composition.

The invention further provides a method of screening for a compound that specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

The invention further provides a method of screening for a compound that modulates the activity of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino

acid sequence selected from the group consisting of SEQ ID NO:1-15, b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

The invention further provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence selected from the group consisting of SEQ ID NO:16-30, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, and b) detecting altered expression of the target polynucleotide.

The invention further provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:16-30, ii) a naturally occurring polynucleotide comprising a polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:16-30, iii) a polynucleotide having a sequence complementary to i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:16-30, ii) a naturally occurring polynucleotide comprising a polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:16-30, iii) a polynucleotide complementary to the polynucleotide of i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide comprises a fragment of a polynucleotide sequence selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of

hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

5 BRIEF DESCRIPTION OF THE TABLES

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the present invention.

Table 2 shows the GenBank identification number and annotation of the nearest GenBank homolog for polypeptides of the invention. The probability score for the match between each
10 polypeptide and its GenBank homolog is also shown.

Table 3 shows structural features of polypeptide sequences of the invention, including predicted motifs and domains, along with the methods, algorithms, and searchable databases used for analysis of the polypeptides.

Table 4 lists the cDNA and genomic DNA fragments which were used to assemble
15 polynucleotide sequences of the invention, along with selected fragments of the polynucleotide sequences.

Table 5 shows the representative cDNA library for polynucleotides of the invention.

Table 6 provides an appendix which describes the tissues and vectors used for construction of the cDNA libraries shown in Table 5.

20 Table 7 shows the tools, programs, and algorithms used to analyze the polynucleotides and polypeptides of the invention, along with applicable descriptions, references, and threshold parameters.

DESCRIPTION OF THE INVENTION

25 Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular machines, materials and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

30 It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

35 Unless defined otherwise, all technical and scientific terms used herein have the same

meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

"TRICH" refers to the amino acid sequences of substantially purified TRICH obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of TRICH. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of TRICH either by directly interacting with TRICH or by acting on components of the biological pathway in which TRICH participates.

An "allelic variant" is an alternative form of the gene encoding TRICH. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding TRICH include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as TRICH or a polypeptide with at least one functional characteristic of TRICH. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding TRICH, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding TRICH. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent TRICH. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of TRICH is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having

similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" refer to an oligopeptide, peptide, polypeptide, or protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid sequence. Amplification is generally carried out using polymerase chain reaction (PCR) technologies well known in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of TRICH. Antagonists may include proteins such as antibodies, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of TRICH either by directly interacting with TRICH or by acting on components of the biological pathway in which TRICH participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, F(ab')₂, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind TRICH polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified

sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" refers to the capability of the natural, recombinant, or synthetic TRICH, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

"Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 3'-TCA-5'.

A "composition comprising a given polynucleotide sequence" and a "composition comprising a given amino acid sequence" refer broadly to any composition containing the given polynucleotide or amino acid sequence. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotide sequences encoding TRICH or fragments of TRICH may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (GCG, Madison WI) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

	Original Residue	Conservative Substitution
	Ala	Gly, Ser
	Arg	His, Lys
	Asn	Asp, Gln, His
5	Asp	Asn, Glu
	Cys	Ala, Ser
	Gln	Asn, Glu, His
	Glu	Asp, Gln, His
	Gly	Ala
10	His	Asn, Arg, Gln, Glu
	Ile	Leu, Val
	Leu	Ile, Val
	Lys	Arg, Gln, Glu
	Met	Leu, Ile
15	Phe	His, Met, Leu, Trp, Tyr
	Ser	Cys, Thr
	Thr	Ser, Val
	Trp	Phe, Tyr
	Tyr	His, Phe, Trp
20	Val	Ile, Leu, Thr

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to a chemically modified polynucleotide or polypeptide. Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

A "fragment" is a unique portion of TRICH or the polynucleotide encoding TRICH which is identical in sequence to but shorter in length than the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from 5 to 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid

residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the

- 5 specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:16-30 comprises a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:16-30, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:16-30 is useful, for
10 example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:16-30 from related polynucleotide sequences. The precise length of a fragment of SEQ ID NO:16-30 and the region of SEQ ID NO:16-30 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-15 is encoded by a fragment of SEQ ID NO:16-30. A fragment
15 of SEQ ID NO:1-15 comprises a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-15. For example, a fragment of SEQ ID NO:1-15 is useful as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-15. The precise length of a fragment of SEQ ID NO:1-15 and the region of SEQ ID NO:1-15 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended
20 purpose for the fragment.

A "full length" polynucleotide sequence is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A "full length" polynucleotide sequence encodes a "full length" polypeptide sequence.

"Homology" refers to sequence similarity or, interchangeably, sequence identity, between
25 two or more polynucleotide sequences or two or more polypeptide sequences.

The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer to the percentage of residue matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and
30 therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in
35 Higgins, D.G. and P.M. Sharp (1989) CABIOS 5:151-153 and in Higgins, D.G. et al. (1992) CABIOS

8:189-191. For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default. Percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polynucleotide sequences.

- 5 Alternatively, a suite of commonly used and freely available sequence comparison algorithms is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at <http://www.ncbi.nlm.nih.gov/BLAST/>. The BLAST software suite includes various sequence
- 10 analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at <http://www.ncbi.nlm.nih.gov/gorf/bl2.html>. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST
- 15 programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Reward for match: 1

- 20 *Penalty for mismatch: -2*

Open Gap: 5 and Extension Gap: 2 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 11

- 25 *Filter: on*

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous

30 nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

- Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes
- 35 in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid

sequences that all encode substantially the same protein.

The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table. As with polynucleotide alignments, the percent identity is reported by CLUSTAL V as the "percent similarity" between aligned polypeptide sequence pairs.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62
Open Gap: 11 and Extension Gap: 1 penalties
Gap x drop-off: 50
Expect: 10
Word Size: 3
Filter: on

Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

"Human artificial chromosomes" (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size and which contain all of the elements required for chromosome replication, segregation and maintenance.

The term "humanized antibody" refers to an antibody molecule in which the amino acid

sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

"Hybridization" refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the "washing" step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 µg/ml sheared, denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating T_m and conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; specifically see volume 2, chapter 9.

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 µg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen bonds between complementary bases. A

hybridization complex may be formed in solution (e.g., C₃t or R₃t analysis) or formed between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

5 The words "insertion" and "addition" refer to changes in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

 "Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect
10 cellular and systemic defense systems.

 An "immunogenic fragment" is a polypeptide or oligopeptide fragment of TRICH which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of TRICH which is useful in any of the antibody production methods disclosed herein or known in the
15 art.

 The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, or other chemical compounds on a substrate.

 The terms "element" and "array element" refer to a polynucleotide, polypeptide, or other chemical compound having a unique and defined position on a microarray.

20 The term "modulate" refers to a change in the activity of TRICH. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of TRICH.

 The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or
25 synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

 "Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding
30 sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

 "Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition.
35 PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript

elongation, and may be pegylated to extend their lifespan in the cell.

"Post-translational modification" of an TRICH may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of TRICH.

"Probe" refers to nucleic acid sequences encoding TRICH, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acid sequences. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes.

"Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in the references, for example Sambrook, J. et al. (1989) Molecular Cloning: A Laboratory Manual, 2nd ed., vol. 1-3, Cold Spring Harbor Press, Plainview NY; Ausubel, F.M. et al. (1987) Current Protocols in Molecular Biology, Greene Publ. Assoc. & Wiley-Intersciences, New York NY; Innis, M. et al. (1990) PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA. PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3

primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection
5 programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both
10 unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

15 A "recombinant nucleic acid" is a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook, supra. The term recombinant includes nucleic acids that have
20 been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a
25 vaccinia virus, that could be used to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription,
30 translation, or RNA stability.

"Reporter molecules" are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

35 An "RNA equivalent," in reference to a DNA sequence, is composed of the same linear

sequence of nucleotides as the reference DNA sequence with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The term "sample" is used in its broadest sense. A sample suspected of containing TRICH, nucleic acids encoding TRICH, or fragments thereof may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least 60% free, preferably at least 75% free, and most preferably at least 90% free from other components with which they are naturally associated.

A "substitution" refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

A "transcript image" refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.

"Transformation" describes a process by which exogenous DNA is introduced into a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed cells" includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. The term genetic manipulation does not include classical cross-breeding, or in vitro fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook et al. (1989), supra.

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternative splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotide sequences that vary from one species to another. The resulting polypeptides will generally have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at

least 60%, at least 70%, at least 80%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length of one of the polypeptides.

5 THE INVENTION

The invention is based on the discovery of new human transporters and ion channels (TRICH), the polynucleotides encoding TRICH, and the use of these compositions for the diagnosis, treatment, or prevention of transport, neurological, muscle, immunological, and cell proliferative disorders.

10 Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide sequences of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an Incyte polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide sequence is
15 denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and an Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown.

Table 2 shows sequences with homology to the polypeptides of the invention as identified by BLAST analysis against the GenBank protein (genpept) database. Columns 1 and 2 show the polypeptide sequence identification number (Polypeptide SEQ ID NO:) and the corresponding Incyte
20 polypeptide sequence number (Incyte Polypeptide ID) for polypeptides of the invention. Column 3 shows the GenBank identification number (Genbank ID NO:) of the nearest GenBank homolog. Column 4 shows the probability score for the match between each polypeptide and its GenBank homolog. Column 5 shows the annotation of the GenBank homolog along with relevant citations where applicable, all of which are expressly incorporated by reference herein.

25 Table 3 shows various structural features of the polypeptides of the invention. Columns 1 and 2 show the polypeptide sequence identification number (SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the number of amino acid residues in each polypeptide. Column 4 shows potential phosphorylation sites, and column 5 shows potential glycosylation sites, as determined by the
30 MOTIFS program of the GCG sequence analysis software package (Genetics Computer Group, Madison WI). Column 6 shows amino acid residues comprising signature sequences, domains, and motifs. Column 7 shows analytical methods for protein structure/function analysis and in some cases, searchable databases to which the analytical methods were applied.

Together, Tables 2 and 3 summarize the properties of polypeptides of the invention, and these
35 properties establish that the claimed polypeptides are transporters and ion channels. For example,

SEQ ID NO:5 is 98% identical, from residue M1 to residue I1009, to *Rattus norvegicus* glutamate receptor delta-1 subunit (GenBank ID g475542) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 0.0, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:5 also contains a receptor family ligand binding region as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein family domains. (See Table 3.) Data from BLIMPS and additional BLAST analyses provide further corroborative evidence that SEQ ID NO:5 is a glutamate receptor. SEQ ID NO:1-4 and SEQ ID NO:6-15 were analyzed and annotated in a similar manner. The algorithms and parameters for the analysis of SEQ ID NO:1-15 are described in Table 7.

As shown in Table 4, the full length polynucleotide sequences of the present invention were assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of these two types of sequences. Columns 1 and 2 list the polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and the corresponding Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) for each polynucleotide of the invention. Column 3 shows the length of each polynucleotide sequence in basepairs. Column 4 lists fragments of the polynucleotide sequences which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:16-30 or that distinguish between SEQ ID NO:16-30 and related polynucleotide sequences. Column 5 shows identification numbers corresponding to cDNA sequences, coding sequences (exons) predicted from genomic DNA, and/or sequence assemblages comprised of both cDNA and genomic DNA. These sequences were used to assemble the full length polynucleotide sequences of the invention. Columns 6 and 7 of Table 4 show the nucleotide start (5') and stop (3') positions of the cDNA and genomic sequences in column 5 relative to their respective full length sequences.

The identification numbers in Column 5 of Table 4 may refer specifically, for example, to Incyte cDNAs along with their corresponding cDNA libraries. For example, 6431853H1 is the identification number of an Incyte cDNA sequence, and LUNGNON07 is the cDNA library from which it is derived. Incyte cDNAs for which cDNA libraries are not indicated were derived from pooled cDNA libraries (e.g., 71456748V1). Alternatively, the identification numbers in column 5 may refer to GenBank cDNAs or ESTs (e.g., g5396013) which contributed to the assembly of the full length polynucleotide sequences. Alternatively, the identification numbers in column 5 may refer to coding regions predicted by Genscan analysis of genomic DNA. For example, GBI.g6742097_000003.fasta.edit is the identification number of a Genscan-predicted coding sequence, with g6742097 being the GenBank identification number of the sequence to which Genscan was applied. The Genscan-predicted coding sequences may have been edited prior to assembly. (See

Example IV.) Alternatively, the identification numbers in column 5 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon stitching" algorithm. (See Example V.) Alternatively, the identification numbers in column 5 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon-stretching" algorithm. (See Example V.) In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in column 5 was obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA identification numbers are not shown.

Table 5 shows the representative cDNA libraries for those full length polynucleotide sequences which were assembled using Incyte cDNA sequences. The representative cDNA library is the Incyte cDNA library which is most frequently represented by the Incyte cDNA sequences which were used to assemble and confirm the above polynucleotide sequences. The tissues and vectors which were used to construct the cDNA libraries shown in Table 5 are described in Table 6.

The invention also encompasses TRICH variants. A preferred TRICH variant is one which has at least about 80%, or alternatively at least about 90%, or even at least about 95% amino acid sequence identity to the TRICH amino acid sequence, and which contains at least one functional or structural characteristic of TRICH.

The invention also encompasses polynucleotides which encode TRICH. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:16-30, which encodes TRICH. The polynucleotide sequences of SEQ ID NO:16-30, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses a variant of a polynucleotide sequence encoding TRICH. In particular, such a variant polynucleotide sequence will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to the polynucleotide sequence encoding TRICH. A particular aspect of the invention encompasses a variant of a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:16-30 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:16-30. Any one of the polynucleotide variants described above can encode an amino acid sequence which contains at least one functional or structural characteristic of TRICH.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding TRICH, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide

sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring TRICH, and all such variations are to be considered as being specifically disclosed.

5 Although nucleotide sequences which encode TRICH and its variants are generally capable of hybridizing to the nucleotide sequence of the naturally occurring TRICH under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding TRICH or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide
10 occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding TRICH and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

15 The invention also encompasses production of DNA sequences which encode TRICH and TRICH derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding TRICH or any fragment thereof.

20 Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NO:16-30 and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G.M. and S.L. Berger (1987) *Methods Enzymol.* 152:399-407; Kimmel, A.R. (1987) *Methods Enzymol.* 152:507-511.) Hybridization conditions, including annealing and wash conditions, are described in
25 "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems), thermostable T7 polymerase (Amersham Pharmacia Biotech, Piscataway NJ), or
30 combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Life Technologies, Gaithersburg MD). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA
35 sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system

(Molecular Dynamics, Sunnyvale CA), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art. (See, e.g., Ausubel, F.M. (1997) Short Protocols in Molecular Biology, John Wiley & Sons, New York NY, unit 7.7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853.)

- 5 The nucleic acid sequences encoding TRICH may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector. (See, e.g., Sarkar, G. (1993) *PCR Methods Applic.* 2:318-322.)
- 10 Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences. (See, e.g., Triglia, T. et al. (1988) *Nucleic Acids Res.* 16:8186.) A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA. (See, e.g., Lagerstrom, M. et al. (1991) *PCR Methods Applic.* 1:111-119.)
- 15 In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art. (See, e.g., Parker, J.D. et al. (1991) *Nucleic Acids Res.* 19:3055-3060.) Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries
- 20 (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 primer analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of
- 25 about 68°C to 72°C.

When screening for full length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence

30 into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate

35

software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

5 In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode TRICH may be cloned in recombinant DNA molecules that direct expression of TRICH, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and used to express TRICH.

10 The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter TRICH-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-
15 mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent Number 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat.
20 Biotechnol. 17:259-264; and Crameri, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of TRICH, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired
25 properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of
30 homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, sequences encoding TRICH may be synthesized, in whole or in part, using chemical methods well known in the art. (See, e.g., Caruthers, M.H. et al. (1980) Nucleic Acids
35 Symp. Ser. 7:215-223; and Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232.)

Alternatively, TRICH itself or a fragment thereof may be synthesized using chemical methods. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques. (See, e.g., Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp. 55-60; and Roberge, J.Y. et al. (1995) *Science* 269:202-204.) Automated synthesis
5 may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence of TRICH, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid
10 chromatography. (See, e.g., Chiez, R.M. and F.Z. Regnier (1990) *Methods Enzymol.* 182:392-421.) The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing. (See, e.g., Creighton, supra, pp. 28-53.)

In order to express a biologically active TRICH, the nucleotide sequences encoding TRICH or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which
15 contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotide sequences encoding TRICH. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of sequences
20 encoding TRICH. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where sequences encoding TRICH and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG
25 initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used. (See, e.g., Scharf, D. et al. (1994) *Results Probl. Cell Differ.* 20:125-162.)

Methods which are well known to those skilled in the art may be used to construct expression
30 vectors containing sequences encoding TRICH and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. (See, e.g., Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview NY, ch. 4, 8, and 16-17; Ausubel, F.M. et al. (1995) Current Protocols in Molecular Biology, John Wiley & Sons, New York NY, ch. 9, 13, and
35 16.)

A variety of expression vector/host systems may be utilized to contain and express sequences encoding TRICH. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus);

5 plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems. (See, e.g., Sambrook, *supra*; Ausubel, *supra*; Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509; Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945; Takamatsu, N. (1987) EMBO

10 J. 6:307-311; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659; and Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355.) Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of nucleotide sequences to the targeted organ, tissue, or cell population. (See, e.g., Di

15 Nicola, M. et al. (1998) Cancer Gen. Ther. 5(6):350-356; Yu, M. et al. (1993) Proc. Natl. Acad. Sci. USA 90(13):6340-6344; Buller, R.M. et al. (1985) Nature 317(6040):813-815; McGregor, D.P. et al. (1994) Mol. Immunol. 31(3):219-226; and Verma, I.M. and N. Somia (1997) Nature 389:239-242.)

The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending

20 upon the use intended for polynucleotide sequences encoding TRICH. For example, routine cloning, subcloning, and propagation of polynucleotide sequences encoding TRICH can be achieved using a multifunctional E. coli vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT 1 plasmid (Life Technologies). Ligation of sequences encoding TRICH into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of

25 transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for in vitro transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence. (See, e.g., Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509.) When large quantities of TRICH are needed, e.g. for the production of antibodies, vectors which direct high level expression of TRICH may be used. For example, vectors

30 containing the strong, inducible SP6 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of TRICH. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast Saccharomyces cerevisiae or Pichia pastoris. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable

35 integration of foreign sequences into the host genome for stable propagation. (See, e.g., Ausubel,

1995, *supra*; Bitter, G.A. et al. (1987) *Methods Enzymol.* 153:516-544; and Scorer, C.A. et al. (1994) *Bio/Technology* 12:181-184.)

Plant systems may also be used for expression of TRICH. Transcription of sequences encoding TRICH may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takanatsu, N. (1987) *EMBO J.* 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used. (See, e.g., Coruzzi, G. et al. (1984) *EMBO J.* 3:1671-1680; Broglie, R. et al. (1984) *Science* 224:838-843; and Winter, J. et al. (1991) *Results Probl. Cell Differ.* 17:85-105.) These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. (See, e.g., The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196.)

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding TRICH may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses TRICH in host cells. (See, e.g., Logan, J. and T. Shenk (1984) *Proc. Natl. Acad. Sci. USA* 81:3655-3659.) In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes. (See, e.g., Harrington, J.J. et al. (1997) *Nat. Genet.* 15:345-355.)

For long term production of recombinant proteins in mammalian systems, stable expression of TRICH in cell lines is preferred. For example, sequences encoding TRICH can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine

phosphoribosyltransferase genes, for use in *trk* and *apr* cells, respectively. (See, e.g., Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823.) Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chloresulfuron and phosphinotricin acetyltransferase, respectively. (See, e.g., Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14.) Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites. (See, e.g., Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051.) Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech), β glucuronidase and its substrate β -glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system. (See, e.g., Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131.)

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding TRICH is inserted within a marker gene sequence, transformed cells containing sequences encoding TRICH can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding TRICH under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the nucleic acid sequence encoding TRICH and that express TRICH may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of TRICH using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on TRICH is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art. (See, e.g., Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; and Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ.)

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding TRICH include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide.

5 Alternatively, the sequences encoding TRICH, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes *in vitro* by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Pharmacia Biotech, Promega
10 (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding TRICH may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein
15 produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode TRICH may be designed to contain signal sequences which direct secretion of TRICH through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the
20 inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for
25 post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding TRICH may be ligated to a heterologous sequence resulting in translation of a
30 fusion protein in any of the aforementioned host systems. For example, a chimeric TRICH protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of TRICH activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST),
35 maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG,

c-myc, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the TRICH encoding sequence and the heterologous protein sequence, so that TRICH may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel (1995, *supra*, ch. 10). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In a further embodiment of the invention, synthesis of radiolabeled TRICH may be achieved *in vitro* using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, ³⁵S-methionine.

TRICH of the present invention or fragments thereof may be used to screen for compounds that specifically bind to TRICH. At least one and up to a plurality of test compounds may be screened for specific binding to TRICH. Examples of test compounds include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

In one embodiment, the compound thus identified is closely related to the natural ligand of TRICH, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner. (See, e.g., Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2): Chapter 5.) Similarly, the compound can be closely related to the natural receptor to which TRICH binds, or to at least a fragment of the receptor, e.g., the ligand binding site. In either case, the compound can be rationally designed using known techniques. In one embodiment, screening for these compounds involves producing appropriate cells which express TRICH, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, *Drosophila*, or *E. coli*. Cells expressing TRICH or cell membrane fractions which contain TRICH are then contacted with a test compound and binding, stimulation, or inhibition of activity of either TRICH or the compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with TRICH, either in solution or affixed to a solid support, and detecting the binding of TRICH to the compound. Alternatively, the assay may detect or measure binding of a test compound in the presence of a

labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

- TRICH of the present invention or fragments thereof may be used to screen for compounds that modulate the activity of TRICH. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for TRICH activity, wherein TRICH is combined with at least one test compound, and the activity of TRICH in the presence of a test compound is compared with the activity of TRICH in the absence of the test compound. A change in the activity of TRICH in the presence of the test compound is indicative of a compound that modulates the activity of TRICH. Alternatively, a test compound is combined with an in vitro or cell-free system comprising TRICH under conditions suitable for TRICH activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of TRICH may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.
- In another embodiment, polynucleotides encoding TRICH or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease. (See, e.g., U.S. Patent Number 5,175,383 and U.S. Patent Number 5,767,337.) For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (neo; Capecchi, M.R. (1989) *Science* 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) *Clin. Invest.* 97:1999-2002; Wagner, K.U. et al. (1997) *Nucleic Acids Res.* 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

- Polynucleotides encoding TRICH may also be manipulated in vitro in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) *Science* 282:1145-1147).

Polynucleotides encoding TRICH can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding TRICH is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress TRICH, e.g., by secreting TRICH in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) *Biotechnol. Annu. Rev.* 4:55-74).

THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of TRICH and transporters and ion channels. In addition, the expression of TRICH is closely associated with brain, prostate and thyroid tissues, neoplasms, and cancers of the small intestine. Therefore, TRICH appears to play a role in transport, neurological, muscle, immunological, and cell proliferative disorders. In the treatment of disorders associated with increased TRICH expression or activity, it is desirable to decrease the expression or activity of TRICH. In the treatment of disorders associated with decreased TRICH expression or activity, it is desirable to increase the expression or activity of TRICH.

Therefore, in one embodiment, TRICH or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of TRICH. Examples of such disorders include, but are not limited to, a transport disorder such as akinesia, amyotrophic lateral sclerosis, ataxia telangiectasia, cystic fibrosis, Becker's muscular dystrophy, Bell's palsy, Charcot-Marie Tooth disease, diabetes mellitus, diabetes insipidus, diabetic neuropathy, Duchenne muscular dystrophy, hyperkalemic periodic paralysis, normokalemic periodic paralysis, Parkinson's disease, malignant hyperthermia, multidrug resistance, myasthenia gravis, myotonic dystrophy, catatonia, tardive dyskinesia, dystonias, peripheral neuropathy, cerebral neoplasms, prostate cancer, cardiac disorders associated with transport, e.g., angina, bradyarrhythmia, tachyarrhythmia, hypertension, Long QT syndrome, myocarditis, cardiomyopathy, nemaline myopathy, centronuclear myopathy, lipid myopathy, mitochondrial myopathy, thyrotoxic myopathy, ethanol myopathy, dermatomyositis, inclusion body myositis, infectious myositis, polymyositis, neurological disorders associated with transport, e.g., Alzheimer's disease, amnesia, bipolar disorder, dementia, depression, epilepsy, Tourette's disorder, paranoid psychoses, and schizophrenia, and other disorders associated with transport, e.g., neurofibromatosis, postherpetic neuralgia, trigeminal neuropathy, sarcoidosis, sickle cell anemia, Wilson's disease, cataracts, infertility, pulmonary artery stenosis, sensorineural autosomal deafness, hyperglycemia, hypoglycemia, Grave's disease, goiter,

- Cushing's disease, Addison's disease, glucose-galactose malabsorption syndrome, hypercholesterolemia, adrenoleukodystrophy, Zellweger syndrome, Menkes disease, occipital horn syndrome, von Gierke disease, cystinuria, iminoglycinuria, Hartup disease, and Fanconi disease; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms,
- 5 Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous
- 10 system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic
- 15 nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathisia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses,
- 20 postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; a muscle disorder such as cardiomyopathy, myocarditis, Duchenne's muscular dystrophy, Becker's muscular dystrophy, myotonic dystrophy, central core disease, nemaline myopathy, centronuclear myopathy, lipid myopathy, mitochondrial myopathy, infectious myositis, polymyositis, dermatomyositis, inclusion body myositis, thyrotoxic
- 25 myopathy, ethanol myopathy, angina, anaphylactic shock, arrhythmias, asthma, cardiovascular shock, Cushing's syndrome, hypertension, hypoglycemia, myocardial infarction, migraine, pheochromocytoma, and myopathies including encephalopathy, epilepsy, Kearns-Sayre syndrome, lactic acidosis, myoclonic disorder, ophthalmoplegia, and acid maltase deficiency (AMD, also known as Pompe's disease); an immunological disorder such as acquired immunodeficiency syndrome
- 30 (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis,

erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, 5 systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; and a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal 10 nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus.

15 In another embodiment, a vector capable of expressing TRICH or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of TRICH including, but not limited to, those described above.

In a further embodiment, a composition comprising a substantially purified TRICH in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent 20 a disorder associated with decreased expression or activity of TRICH including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of TRICH may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of TRICH including, but not limited to, those listed above.

25 In a further embodiment, an antagonist of TRICH may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of TRICH. Examples of such disorders include, but are not limited to, those transport, neurological, muscle, immunological, and cell proliferative disorders described above. In one aspect, an antibody which specifically binds TRICH may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for 30 bringing a pharmaceutical agent to cells or tissues which express TRICH.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding TRICH may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of TRICH including, but not limited to, those described above.

In other embodiments, any of the proteins, antagonists, antibodies, agonists, complementary

sequences, or vectors of the invention may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of TRICH may be produced using methods which are generally known in the art. In particular, purified TRICH may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind TRICH. Antibodies to TRICH may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies (i.e., those which inhibit dimer formation) are generally preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others may be immunized by injection with TRICH or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and *Corynebacterium parvum* are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to TRICH have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are identical to a portion of the amino acid sequence of the natural protein. Short stretches of TRICH amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to TRICH may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique. (See, e.g., Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; and Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120.)

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used. (See, e.g., Morrison, S.L. et al. (1984) Proc.

Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; and Takeda, S. et al. (1985) Nature 314:452-454.) Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce TRICH-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries. (See, e.g., Burton, D.R. (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137.)

Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature. (See, e.g., Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299.)

Antibody fragments which contain specific binding sites for TRICH may also be generated. For example, such fragments include, but are not limited to, $F(ab)_2$ fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the $F(ab)_2$ fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity. (See, e.g., Huse, W.D. et al. (1989) Science 246:1275-1281.)

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between TRICH and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering TRICH epitopes is generally used, but a competitive binding assay may also be employed (Pound, supra).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for TRICH. Affinity is expressed as an association constant, K_a , which is defined as the molar concentration of TRICH-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The K_a determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple TRICH epitopes, represents the average affinity, or avidity, of the antibodies for TRICH. The K_a determined for a preparation of monoclonal antibodies, which are monospecific for a particular TRICH epitope, represents a true measure of affinity. High-affinity antibody preparations with K_a ranging from about 10^9 to 10^{12} L/mole are preferred for use in immunoassays in which the TRICH-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with K_a ranging from about 10^6 to 10^7 L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of TRICH,

preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to
 5 determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of TRICH-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available. (See, e.g.,
 10 Catty, supra, and Coligan et al. supra.)

In another embodiment of the invention, the polynucleotides encoding TRICH, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene
 15 encoding TRICH. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding TRICH. (See, e.g., Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press Inc., Totawa NJ.)

In therapeutic use, any gene delivery system suitable for introduction of the antisense
 20 sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein. (See, e.g., Slater, J.E. et al. (1998) *J. Allergy Clin. Immunol.* 102(3):469-475; and Scanlon, K.J. et al. (1995) 9(13):1288-1296.) Antisense sequences can also be introduced intracellularly through the use of viral
 25 vectors, such as retrovirus and adeno-associated virus vectors. (See, e.g., Miller, A.D. (1990) *Blood* 76:271; Ausubel, supra; Uckert, W. and W. Walther (1994) *Pharmacol. Ther.* 63(3):323-347.) Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art. (See, e.g., Rossi, J.J. (1995) *Br. Med. Bull.* 51(1):217-225; Boado, R.J. et al. (1998) *J. Pharm. Sci.* 87(11):1308-1315; and Morris, M.C. et al. (1997) *Nucleic Acids Res.*
 30 25(14):2730-2736.)

In another embodiment of the invention, polynucleotides encoding TRICH may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) *Science* 288:669-672), severe combined
 35 immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency

(Blaese, R.M. et al. (1995) Science 270:475-480; Bordignon, C. et al. (1995) Science 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) Cell 75:207-216; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:643-666; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:667-703), thalassemias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) Science 270:404-410; Verma, I.M. and N. Somia (1997) Nature 389:239-242)), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) Nature 335:395-396; Poeschla, E. et al. (1996) Proc. Natl. Acad. Sci. USA. 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as Candida albicans and Paracoccidioides brasiliensis; and protozoan parasites such as Plasmodium falciparum and Trypanosoma cruzi). In the case where a genetic deficiency in TRICH expression or regulation causes disease, the expression of TRICH from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in TRICH are treated by constructing mammalian expression vectors encoding TRICH and introducing these vectors by mechanical means into TRICH-deficient cells. Mechanical transfer technologies for use with cells in vivo or ex vitro include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) Annu. Rev. Biochem. 62:191-217; Ivics, Z. (1997) Cell 91:501-510; Boulay, J.-L. and H. Récipon (1998) Curr. Opin. Biotechnol. 9:445-450).

Expression vectors that may be effective for the expression of TRICH include, but are not limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). TRICH may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or β -actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) Proc. Natl. Acad. Sci. USA 89:5547-5551; Gossen, M. et al. (1995) Science 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and Blau, H.M. supra), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding TRICH from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) *Virology* 52:456-467), or by electroporation (Neumann, E. et al. (1982) *EMBO J.* 1:841-845). The introduction of DNA to primary cells requires modification of these standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to TRICH expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding TRICH under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) *Proc. Natl. Acad. Sci. USA* 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) *J. Virol.* 61:1647-1650; Bender, M.A. et al. (1987) *J. Virol.* 61:1639-1646; Adam, M.A. and A.D. Miller (1988) *J. Virol.* 62:3802-3806; Dull, T. et al. (1998) *J. Virol.* 72:8463-8471; Zufferey, R. et al. (1998) *J. Virol.* 72:9873-9880). U.S. Patent Number 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., CD4⁺ T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) *J. Virol.* 71:7020-7029; Bauer, G. et al. (1997) *Blood* 89:2259-2267; Bonyhadi, M.L. (1997) *J. Virol.* 71:4707-4716; Ranga, U. et al. (1998) *Proc. Natl. Acad. Sci. USA* 95:1201-1206; Su, L. (1997) *Blood* 89:2283-2290).

In the alternative, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding TRICH to cells which have one or more genetic abnormalities with respect to the expression of TRICH. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Cssete, M.E. et al. (1995) *Transplantation* 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent Number 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"),

hereby incorporated by reference. For adenoviral vectors, see also Antinuzzi, P.A. et al. (1999) *Annu. Rev. Nutr.* 19:511-544 and Verma, I.M. and N. Somia (1997) *Nature* 18:389:239-242, both incorporated by reference herein.

In another alternative, a herpes-based, gene therapy delivery system is used to deliver
5 polynucleotides encoding TRICH to target cells which have one or more genetic abnormalities with respect to the expression of TRICH. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing TRICH to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has
10 been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) *Exp. Eye Res.* 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent Number 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent Number 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a
15 cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999) *J. Virol.* 73:519-532 and Xu, H. et al. (1994) *Dev. Biol.* 163:152-161, hereby incorporated by reference. The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple
20 plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

In another alternative, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding TRICH to target cells. The biology of the prototypic alphavirus,
25 Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Garoff, H. and K.-J. Li (1998) *Curr. Opin. Biotechnol.* 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity
30 (e.g., protease and polymerase). Similarly, inserting the coding sequence for TRICH into the alphavirus genome in place of the capsid-coding region results in the production of a large number of TRICH-coding RNAs and the synthesis of high levels of TRICH in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN)
35 indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy

application (Dryga, S.A. et al. (1997) Virology 228:74-83). The wide host range of alphaviruses will allow the introduction of TRICH into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA

5 transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful
10 because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature. (See, e.g., Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177.) A complementary sequence or antisense molecule may also be designed to block translation of
15 mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze
20 endonucleolytic cleavage of sequences encoding TRICH.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for
25 secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques
30 for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding TRICH. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into
35 cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding TRICH. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased TRICH expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding TRICH may be therapeutically useful, and in the treatment of disorders associated with decreased TRICH expression or activity, a compound which specifically promotes expression of the polynucleotide encoding TRICH may be therapeutically useful.

At least one, and up to a plurality, of test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding TRICH is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an *in vitro* cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding TRICH are assayed by any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding TRICH. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without

exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a *Schizosaccharomyces pombe* gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruce, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruce, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable for use *in vivo*, *in vitro*, and *ex vivo*. For *ex vivo* therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art. (See, e.g., Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466.)

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various formulations are commonly known and are thoroughly discussed in the latest edition of *Remington's Pharmaceutical Sciences* (Maack Publishing, Easton PA). Such compositions may consist of TRICH, antibodies to TRICH, and mimetics, agonists, antagonists, or inhibitors of TRICH.

The compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

Compositions for pulmonary administration may be prepared in liquid or dry powder form. These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the

lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery has the advantage of administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

Specialized forms of compositions may be prepared for direct intracellular delivery of macromolecules comprising TRICH or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, TRICH or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) *Science* 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example TRICH or fragments thereof, antibodies of TRICH, and agonists, antagonists or inhibitors of TRICH, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED_{50} (the dose therapeutically effective in 50% of the population) or LD_{50} (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the LD_{50}/ED_{50} ratio. Compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED_{50} with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and

response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 μg to 100,000 μg , up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

DIAGNOSTICS

10 In another embodiment, antibodies which specifically bind TRICH may be used for the diagnosis of disorders characterized by expression of TRICH, or in assays to monitor patients being treated with TRICH or agonists, antagonists, or inhibitors of TRICH. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for TRICH include methods which utilize the antibody and a label to detect TRICH
15 in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring TRICH, including ELISAs, RIAs, and FACS, are known
20 in the art and provide a basis for diagnosing altered or abnormal levels of TRICH expression. Normal or standard values for TRICH expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to TRICH under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of TRICH expressed in
25 subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding TRICH may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect
30 and quantify gene expression in biopsied tissues in which expression of TRICH may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of TRICH, and to monitor regulation of TRICH levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding TRICH or closely related molecules may be used
35 to identify nucleic acid sequences which encode TRICH. The specificity of the probe, whether it is

made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding TRICH, allelic variants, or related sequences.

- 5 Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the TRICH encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO: 16-30 or from genomic sequences including promoters, enhancers, and introns of the TRICH gene.

- Means for producing specific hybridization probes for DNAs encoding TRICH include the
10 cloning of polynucleotide sequences encoding TRICH or TRICH derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as ^{32}P or ^{35}S , or by enzymatic labels,
15 such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

- Polynucleotide sequences encoding TRICH may be used for the diagnosis of disorders associated with expression of TRICH. Examples of such disorders include, but are not limited to, a transport disorder such as akinesia, amyotrophic lateral sclerosis, ataxia telangiectasia, cystic fibrosis, Becker's muscular dystrophy, Bell's palsy, Charcot-Marie Tooth disease, diabetes mellitus, diabetes
20 insipidus, diabetic neuropathy, Duchenne muscular dystrophy, hyperkalemic periodic paralysis, normokalemic periodic paralysis, Parkinson's disease, malignant hyperthermia, multidrug resistance, myasthenia gravis, myotonic dystrophy, catatonia, tardive dyskinesia, dystonias, peripheral neuropathy, cerebral neoplasms, prostate cancer, cardiac disorders associated with transport, e.g., angina, bradyarrhythmia, tachyarrhythmia, hypertension, Long QT syndrome, myocarditis,
25 cardiomyopathy, nemaline myopathy, centronuclear myopathy, lipid myopathy, mitochondrial myopathy, thyrotoxic myopathy, ethanol myopathy, dermatomyositis, inclusion body myositis, infectious myositis, polymyositis, neurological disorders associated with transport, e.g., Alzheimer's disease, amnesia, bipolar disorder, dementia, depression, epilepsy, Tourette's disorder, paranoid psychoses, and schizophrenia, and other disorders associated with transport, e.g., neurofibromatosis,
30 postherpetic neuralgia, trigeminal neuropathy, sarcoidosis, sickle cell anemia, Wilson's disease, cataracts, infertility, pulmonary artery stenosis, sensorineural autosomal deafness, hyperglycemia, hypoglycemia, Grave's disease, goiter, Cushing's disease, Addison's disease, glucose-galactose malabsorption syndrome, hypercholesterolemia, adrenoleukodystrophy, Zellweger syndrome, Menkes disease, occipital horn syndrome, von Gierke disease, cystinuria, iminoglycinuria, Hartup
35 disease, and Fanconi disease; a neurological disorder such as epilepsy, ischemic cerebrovascular

disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathisia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; a muscle disorder such as cardiomyopathy, myocarditis, Duchenne's muscular dystrophy, Becker's muscular dystrophy, myotonic dystrophy, central core disease, nemaline myopathy, centronuclear myopathy, lipid myopathy, mitochondrial myopathy, infectious myositis, polymyositis, dermatomyositis, inclusion body myositis, thyrotoxic myopathy, ethanol myopathy, angina, anaphylactic shock, arrhythmias, asthma, cardiovascular shock, Cushing's syndrome, hypertension, hypoglycemia, myocardial infarction, migraine, pheochromocytoma, and myopathies including encephalopathy, epilepsy, Kearns-Sayre syndrome, lactic acidosis, myoclonic disorder, ophthalmoplegia, and acid maltase deficiency (AMD, also known as Pompe's disease); an immunological disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis,

psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; and a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus. The polynucleotide sequences encoding TRICH may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered TRICH expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding TRICH may be useful in assays that detect the presence of associated disorders, particularly those mentioned above. The nucleotide sequences encoding TRICH may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of nucleotide sequences encoding TRICH in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of TRICH, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding TRICH, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated,

hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

- 5 With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier thereby preventing the development
10 or further progression of the cancer.

- Additional diagnostic uses for oligonucleotides designed from the sequences encoding TRICH may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced *in vitro*. Oligomers will preferably contain a fragment of a polynucleotide encoding TRICH, or a fragment of a polynucleotide complementary to the polynucleotide encoding
15 TRICH, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

- In a particular aspect, oligonucleotide primers derived from the polynucleotide sequences encoding TRICH may be used to detect single nucleotide polymorphisms (SNPs). SNPs are
20 substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from the polynucleotide sequences encoding TRICH are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example,
25 from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSSCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis
30 methods, termed *in silico* SNP (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry
35 using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

Methods which may also be used to quantify the expression of TRICH include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves. (See, e.g., Melby, P.C. et al. (1993) *J. Immunol. Methods* 159:235-244; Duplaa, C. et al. (1993) *Anal. Biochem.* 212:229-236.) The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotide sequences described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described below. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

In another embodiment, TRICH, fragments of TRICH, or antibodies specific for TRICH may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a given time. (See Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent Number 5,840,484, expressly incorporated by reference herein.) Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression in vivo,

as in the case of a tissue or biopsy sample, or in vitro, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with in vitro model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) *Mol. Carcinog.* 24:153-159; Steiner, S. and N.L. Anderson (2000) *Toxicol. Lett.* 112-113:467-471, expressly incorporated by reference herein). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity. (See, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at <http://www.niehs.nih.gov/oc/news/toxchip.htm>.) Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

In one embodiment, the toxicity of a test compound is assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared with levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another particular embodiment relates to the use of the polypeptide sequences of the present invention to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the

separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, supra). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of the present invention. In some cases, further sequence data may be obtained for definitive protein identification.

A proteomic profile may also be generated using antibodies specific for TRICH to quantify the levels of TRICH expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) *Anal. Biochem.* 270:103-111; Mendoz, L.G. et al. (1999) *Biotechniques* 27:773-788). Detection may be performed by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Seilhamer (1997) *Electrophoresis* 18:533-537), so proteome toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic

response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the polypeptides of the present invention.

- In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art. (See, e.g., Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/251116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; and Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662.) Various types of microarrays are well known and thoroughly described in DNA Microarrays: A Practical Approach, M. Schena, ed. (1999) Oxford University Press, London, hereby expressly incorporated by reference.

- In another embodiment of the invention, nucleic acid sequences encoding TRICH may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries. (See, e.g., Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; and Trask, B.J. (1991) Trends Genet. 7:149-154.) Once mapped, the nucleic acid sequences of the invention may be used to develop genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP). (See, for example, Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357.)

- Fluorescent in situ hybridization (FISH) may be correlated with other physical and genetic map data. (See, e.g., Heinz-Ulrich, et al. (1995) in Meyers, supra, pp. 965-968.) Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man

(OMIM) World Wide Web site. Correlation between the location of the gene encoding TRICH on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as
5 linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely
10 localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation. (See, e.g., Gatti, R.A. et al. (1988) Nature 336:577-580.) The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

15 In another embodiment of the invention, TRICH, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between TRICH and the agent being tested may be measured.

20 Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest. (See, e.g., Geysen, et al. (1984) PCT application WO84/03564.) In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with TRICH, or fragments thereof, and washed. Bound TRICH is then detected by methods well known in the art. Purified TRICH can
25 also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding TRICH specifically compete with a test compound for binding TRICH.
30 In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with TRICH.

In additional embodiments, the nucleotide sequences which encode TRICH may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such
35 properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

- 5 The disclosures of all patents, applications and publications, mentioned above and below, in particular U.S. Ser. No. 60/195,595, U.S. Ser. No. 60/196,872, U.S. Ser. No. 60/199,020, U.S. Ser. No. 60/200,552, U.S. Ser. No. 60/202,348, and U.S. Ser. No. 60/203,495, are expressly incorporated by reference herein.

10 EXAMPLES

I. Construction of cDNA Libraries

- Incute cDNAs were derived from cDNA libraries described in the LIFESEQ GOLD database (Incute Genomics, Palo Alto CA) and shown in Table 4, column 5. Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a
15 suitable mixture of denaturants, such as TRIZOL (Life Technologies), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

- Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA
20 purity. In some cases, RNA was treated with DNase. For most libraries, poly(A)⁺ RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)⁺PURE mRNA purification kit (Ambion, Austin TX).

- 25 In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERSCRIP^T plasmid system (Life Technologies), using the recommended procedures or similar methods known in the art. (See, e.g., Ausubel, 1997, supra, units 5.1-6.6.) Reverse transcription was initiated using oligo d(T) or random primers. Synthetic
30 oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Pharmacia Biotech) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g.,
35 PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Life Technologies), PCDNA2.1 plasmid

(Invitrogen, Carlsbad CA), PBK-CMV plasmid (Stratagene), or pINCY (Incyte Genomics, Palo Alto CA), or derivatives thereof. Recombinant plasmids were transformed into competent E. coli cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5 α , DH10B, or ElectroMAX DH10B from Life Technologies.

5

II. Isolation of cDNA Clones

Plasmids obtained as described in Example I were recovered from host cells by in vivo excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an
5 AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4 °C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a
10 high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

15 III. Sequencing and Analysis

Incute cDNA recovered in plasmids as described in Example II were sequenced as follows. Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the
20 MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Pharmacia Biotech or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Molecular Dynamics); the
25 ABI PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading frames within the cDNA sequences were identified using standard methods (reviewed in Ausubel, 1997, supra, unit 7.7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example VIII.

The polynucleotide sequences derived from Incute cDNAs were validated by removing vector, linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The Incute cDNA sequences or translations thereof were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and
35 BLOCKS, PRINTS, DOMO, PRODOM, and hidden Markov model (HMM)-based protein family

databases such as PFAM. (HMM is a probabilistic approach which analyzes consensus primary structures of gene families. See, for example, Eddy, S.R. (1996) *Curr. Opin. Struct. Biol.* 6:361-365.) The queries were performed using programs based on BLAST, FASTA, BLIMPS, and HMMER. The Incyte cDNA sequences were assembled to produce full length polynucleotide sequences.

- 5 Alternatively, GenBank cDNAs, GenBank ESTs, stitched sequences, stretched sequences, or Genscan-predicted coding sequences (see Examples IV and V) were used to extend Incyte cDNA assemblages to full length. Assembly was performed using programs based on Phred, Phrap, and Consed, and cDNA assemblages were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive
- 10 the corresponding full length polypeptide sequences. Alternatively, a polypeptide of the invention may begin at any of the methionine residues of the full length translated polypeptide. Full length polypeptide sequences were subsequently analyzed by querying against databases such as the GenBank protein databases (genpept), SwissProt, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, and hidden Markov model (HMM)-based protein family databases such as PFAM. Full length
- 15 polynucleotide sequences are also analyzed using MACDNASIS PRO software (Hitachi Software Engineering, South San Francisco CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments are generated using default parameters specified by the CLUSTAL algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

- 20 Table 7 summarizes the tools, programs, and algorithms used for the analysis and assembly of Incyte cDNA and full length sequences and provides applicable descriptions, references, and threshold parameters. The first column of Table 7 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column
- 25 presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score or the lower the probability value, the greater the identity between two sequences).

- The programs described above for the assembly and analysis of full length polynucleotide and polypeptide sequences were also used to identify polynucleotide sequence fragments from SEQ
- 30 ID NO:16-30. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 4.

IV. Identification and Editing of Coding Sequences from Genomic DNA

- Putative transporters and ion channels were initially identified by running the Genscan gene identification program against public genomic sequence databases (e.g., gbpi and gbhtg). Genscan is
- 35 a general-purpose gene identification program which analyzes genomic DNA sequences from a

variety of organisms (See Burge, C. and S. Karlin (1997) *J. Mol. Biol.* 268:78-94, and Burge, C. and S. Karlin (1998) *Curr. Opin. Struct. Biol.* 8:346-354). The program concatenates predicted exons to form an assembled cDNA sequence extending from a methionine to a stop codon. The output of Genscan is a FASTA database of polynucleotide and polypeptide sequences. The maximum range of
5 Genscan for Genscan to analyze at once was set to 30 kb. To determine which of these Genscan predicted cDNA sequences encode transporters and ion channels, the encoded polypeptides were analyzed by querying against PFAM models for transporters and ion channels. Potential transporters and ion channels were also identified by homology to Incyte cDNA sequences that had been annotated as transporters and ion channels. These selected Genscan-predicted sequences were then
10 compared by BLAST analysis to the genpept and gbprl public databases. Where necessary, the Genscan-predicted sequences were then edited by comparison to the top BLAST hit from genpept to correct errors in the sequence predicted by Genscan, such as extra or omitted exons. BLAST analysis was also used to find any Incyte cDNA or public cDNA coverage of the Genscan-predicted sequences, thus providing evidence for transcription. When Incyte cDNA coverage was available,
15 this information was used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences were obtained by assembling Genscan-predicted coding sequences with Incyte cDNA sequences and/or public cDNA sequences using the assembly process described in Example III. Alternatively, full length polynucleotide sequences were derived entirely from edited or unedited Genscan-predicted coding sequences.

20 **V. Assembly of Genomic Sequence Data with cDNA Sequence Data**
"Stitched" Sequences

Partial cDNA sequences were extended with exons predicted by the Genscan gene identification program described in Example IV. Partial cDNAs assembled as described in Example III were mapped to genomic DNA and parsed into clusters containing related cDNAs and Genscan
25 exon predictions from one or more genomic sequences. Each cluster was analyzed using an algorithm based on graph theory and dynamic programming to integrate cDNA and genomic information, generating possible splice variants that were subsequently confirmed, edited, or extended to create a full length sequence. Sequence intervals in which the entire length of the interval was present on more than one sequence in the cluster were identified, and intervals thus identified were considered to
30 be equivalent by transitivity. For example, if an interval was present on a cDNA and two genomic sequences, then all three intervals were considered to be equivalent. This process allows unrelated but consecutive genomic sequences to be brought together, bridged by cDNA sequence. Intervals thus identified were then "stitched" together by the stitching algorithm in the order that they appear along their parent sequences to generate the longest possible sequence, as well as sequence variants.
35 Linkages between intervals which proceed along one type of parent sequence (cDNA to cDNA or

genomic sequence to genomic sequence) were given preference over linkages which change parent type (cDNA to genomic sequence). The resultant stitched sequences were translated and compared by BLAST analysis to the genpept and gbpi public databases. Incorrect exons predicted by Genscan were corrected by comparison to the top BLAST hit from genpept. Sequences were further extended

- 5 with additional cDNA sequences, or by inspection of genomic DNA, when necessary.

"Stretched" Sequences

- Partial DNA sequences were extended to full length with an algorithm based on BLAST analysis. First, partial cDNAs assembled as described in Example III were queried against public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases
- 10 using the BLAST program. The nearest GenBank protein homolog was then compared by BLAST analysis to either Incyte cDNA sequences or GenScan exon predicted sequences described in Example IV. A chimeric protein was generated by using the resultant high-scoring segment pairs (HSPs) to map the translated sequences onto the GenBank protein homolog. Insertions or deletions may occur in the chimeric protein with respect to the original GenBank protein homolog. The
- 15 GenBank protein homolog, the chimeric protein, or both were used as probes to search for homologous genomic sequences from the public human genome databases. Partial DNA sequences were therefore "stretched" or extended by the addition of homologous genomic sequences. The resultant stretched sequences were examined to determine whether it contained a complete gene.

VI. Chromosomal Mapping of TRICH Encoding Polynucleotides

- 20 The sequences which were used to assemble SEQ ID NO:16-30 were compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:16-30 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available
- 25 from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO:, to that map location.

- Map locations are represented by ranges, or intervals, of human chromosomes. The map
- 30 position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation
- 35 hybrid markers whose sequences were included in each of the clusters. Human genome maps and

other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site (<http://www.ncbi.nlm.nih.gov/genemap/>), can be employed to determine if previously identified disease genes map within or in proximity to the intervals indicated above.

VII. Analysis of Polynucleotide Expression

- 5 Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound. (See, e.g., Sambrook, *supra*, ch. 7; Ausubel (1995) *supra*, ch. 4 and 16.)

- 10 Analogous computer techniques applying BLAST were used to search for identical or related molecules in cDNA databases such as GenBank or LIFESEQ (Incyte Genomics). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$15 \quad \frac{\text{BLAST Score} \times \text{Percent Identity}}{5 \times \text{minimum \{length(Seq. 1), length(Seq. 2)\}}}$$

- The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is
20 calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate
25 the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79%
30 identity and 100% overlap.

- Alternatively, polynucleotide sequences encoding TRICH are analyzed with respect to the tissue sources from which they were derived. For example, some full length sequences are assembled, at least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is
35 classified into one of the following organ/tissue categories: cardiovascular system; connective tissue;

digestive system; embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ cells; hemic and immune system; liver; musculoskeletal system; nervous system; pancreas; respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed; or urinary tract. The number of libraries in each category is counted and divided by the total number of libraries across all categories. Similarly, each human tissue is classified into one of the following disease/condition categories: cancer, cell line, developmental, inflammation, neurological, trauma, cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided by the total number of libraries across all categories. The resulting percentages reflect the tissue- and disease-specific expression of cDNA encoding TRICH. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA).

VIII. Extension of TRICH Encoding Polynucleotides

Full length polynucleotide sequences were also produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer was synthesized to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg^{2+} , $(NH_4)_2SO_4$, and 2-mercaptoethanol, Taq DNA polymerase (Amersham Pharmacia Biotech), ELONGASE enzyme (Life Technologies), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100 μ l PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 μ l of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II

(Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 μ l to 10 μ l aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose gel to determine which reactions were successful in extending the sequence.

- 5 The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Pharmacia Biotech). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones
10 were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Pharmacia Biotech), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent *E. coli* cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.
- 15 The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Pharmacia Biotech) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA
20 recoveries were reamplified using the same conditions as described above. Samples were diluted with 20% dimethylsulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Pharmacia Biotech) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

- In like manner, full length polynucleotide sequences are verified using the above procedure or
25 are used to obtain 5' regulatory sequences using the above procedure along with oligonucleotides designed for such extension, and an appropriate genomic library.

IX. Labeling and Use of Individual Hybridization Probes

- Hybridization probes derived from SEQ ID NO:16-30 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base
30 pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μ Ci of [γ -³²P] adenosine triphosphate (Amersham Pharmacia Biotech), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a
35 SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Pharmacia Biotech).

An aliquot containing 10^7 counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

- The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

10 X. Microarrays

- The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing, See, e.g., Baldeschweiler, supra), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Schna (1999), supra). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements. (See, e.g., Schna, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645; Marshall, A. and J. Hodgson (1998) Nat. Biotechnol. 16:27-31.)

- Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorption and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

Tissue or Cell Sample Preparation

- Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)⁺ RNA is purified using the oligo-(dT) cellulose method. Each poly(A)⁺ RNA sample is

- reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/ μ l oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/ μ l RNase inhibitor, 500 μ M dATP, 500 μ M dGTP, 500 μ M dTTP, 40 μ M dCTP, 40 μ M dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Pharmacia Biotech). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)⁺ RNA with
- 5 GEMBRIGHT kits (Incyte). Specific control poly(A)⁺ RNAs are synthesized by in vitro transcription from non-coding yeast genomic DNA. After incubation at 37°C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85°C to the stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (CLONTECH Laboratories, Inc.
- 10 (CLONTECH), Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14 μ l 5X SSC/0.2% SDS.

Microarray Preparation

- 15 Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5 μ g. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Pharmacia
- 20 Biotech).

- Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water,
- 25 and coated with 0.05% aminopropyl silane (Sigma) in 95% ethanol. Coated slides are cured in a 110°C oven.

- Array elements are applied to the coated glass substrate using a procedure described in US Patent No. 5,807,522, incorporated herein by reference. 1 μ l of the array element DNA, at an average concentration of 100 ng/ μ l, is loaded into the open capillary printing element by a high-speed robotic
- 30 apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene). Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60°C followed by washes in

0.2% SDS and distilled water as before.

Hybridization

Hybridization reactions contain 9 μ l of sample mixture consisting of 0.2 μ g each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65°C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm² coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140 μ l of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60°C. The arrays are washed for 10 min at 45°C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45°C in a second wash buffer (0.1X SSC), and dried.

Detection

Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte).

XI. Complementary Polynucleotides

Sequences complementary to the TRICH-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring TRICH. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of TRICH. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the TRICH-encoding transcript.

XII. Expression of TRICH

Expression and purification of TRICH is achieved using bacterial or virus-based expression systems. For expression of TRICH in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express TRICH upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of TRICH in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant *Autographica californica* nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding TRICH by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to

infect Spodoptera frugiperda (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus. (See Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945.)

- 5 In most expression systems, TRICH is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from Schistosoma japonicum, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Pharmacia Biotech). Following purification, the GST moiety can be proteolytically cleaved from TRICH at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel (1995, supra, ch. 10 and 16). Purified TRICH obtained by these methods can be used directly in the assays shown in Examples XVI, XVII, and XVIII, where applicable.

XIII. Functional Assays

- TRICH function is assessed by expressing the sequences encoding TRICH at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include PCMV SPORT (Life Technologies) and PCR3.1 (Invitrogen, Carlsbad CA), both of which contain the cytomegalovirus promoter. 5-10 μ g of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2 μ g of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with

specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994) Flow Cytometry, Oxford, New York NY.

The influence of TRICH on gene expression can be assessed using highly purified
5 populations of cells transfected with sequences encoding TRICH and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in
10 the art. Expression of mRNA encoding TRICH and other genes of interest can be analyzed by northern analysis or microarray techniques.

XIV. Production of TRICH Specific Antibodies

TRICH substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) *Methods Enzymol.* 182:488-495), or other purification techniques, is used to
15 immunize rabbits and to produce antibodies using standard protocols.

Alternatively, the TRICH amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well
20 described in the art. (See, e.g., Ausubel, 1995, supra, ch. 11.)

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Applied Biosystems) using Fmoc chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity. (See, e.g., Ausubel, 1995, supra.) Rabbits are immunized with the
25 oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-TRICH activity by, for example, binding the peptide or TRICH to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

XV. Purification of Naturally Occurring TRICH Using Specific Antibodies

Naturally occurring or recombinant TRICH is substantially purified by immunoaffinity chromatography using antibodies specific for TRICH. An immunoaffinity column is constructed by covalently coupling anti-TRICH antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Pharmacia Biotech). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

35 Media containing TRICH are passed over the immunoaffinity column, and the column is

washed under conditions that allow the preferential absorbance of TRICH (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/TRICH binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and TRICH is collected.

5 XVI. Identification of Molecules Which Interact with TRICH

Molecules which interact with TRICH may include transporter substrates, agonists or antagonists, modulatory proteins such as G $\beta\gamma$ proteins (Reimann, *supra*) or proteins involved in TRICH localization or clustering such as MAGUKs (Craven, *supra*). TRICH, or biologically active fragments thereof, are labeled with ¹²⁵I Bolton-Hunter reagent. (See, e.g., Bolton A.E. and W.M. Hunter (1973) *Biochem. J.* 133:529-539.) Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled TRICH, washed, and any wells with labeled TRICH complex are assayed. Data obtained using different concentrations of TRICH are used to calculate values for the number, affinity, and association of TRICH with the candidate molecules.

Alternatively, proteins that interact with TRICH are isolated using the yeast 2-hybrid system (Fields, S. and O. Song (1989) *Nature* 340:245-246). TRICH, or fragments thereof, are expressed as fusion proteins with the DNA binding domain of Gal4 or IexA, and potential interacting proteins are expressed as fusion proteins with an activation domain. Interactions between the TRICH fusion protein and the TRICH interacting proteins (fusion proteins with an activation domain) reconstitute a transactivation function that is observed by expression of a reporter gene. Yeast 2-hybrid systems are commercially available, and methods for use of the yeast 2-hybrid system with ion channel proteins are discussed in Niethammer, M. and M. Sheng (1998, *Meth. Enzymol.* 293:104-122).

TRICH may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT) which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

Potential TRICH agonists or antagonists may be tested for activation or inhibition of TRICH ion channel activity using the assays described in section XVIII.

XVII. Demonstration of TRICH Activity

Ion channel activity of TRICH is demonstrated using an electrophysiological assay for ion conductance. TRICH can be expressed by transforming a mammalian cell line such as COS7, HeLa or CHO with a eukaryotic expression vector encoding TRICH. Eukaryotic expression vectors are commercially available, and the techniques to introduce them into cells are well known to those skilled in the art. A second plasmid which expresses any one of a number of marker genes, such as β -galactosidase, is co-transformed into the cells to allow rapid identification of those cells which have taken up and expressed the foreign DNA. The cells are incubated for 48-72 hours after

transformation under conditions appropriate for the cell line to allow expression and accumulation of TRICH and β -galactosidase.

Transformed cells expressing β -galactosidase are stained blue when a suitable colorimetric substrate is added to the culture media under conditions that are well known in the art. Stained cells are tested for differences in membrane conductance by electrophysiological techniques that are well known in the art. Untransformed cells, and/or cells transformed with either vector sequences alone or β -galactosidase sequences alone, are used as controls and tested in parallel. Cells expressing TRICH will have higher anion or cation conductance relative to control cells. The contribution of TRICH to conductance can be confirmed by incubating the cells using antibodies specific for TRICH. The antibodies will bind to the extracellular side of TRICH, thereby blocking the pore in the ion channel, and the associated conductance.

Alternatively, ion channel activity of TRICH is measured as current flow across a TRICH-containing *Xenopus laevis* oocyte membrane using the two-electrode voltage-clamp technique (Ishi et al., *supra*; Jegla, T. and L. Salkoff (1997) J. Neurosci. 17:32-44). TRICH is subcloned into an appropriate *Xenopus* oocyte expression vector, such as pBF, and 0.5-5 ng of mRNA is injected into mature stage IV oocytes. Injected oocytes are incubated at 18 °C for 1-5 days. Inside-out macropatches are excised into an intracellular solution containing 116 mM K-gluconate, 4 mM KCl, and 10 mM Hepes (pH 7.2). The intracellular solution is supplemented with varying concentrations of the TRICH mediator, such as cAMP, cGMP, or Ca^{2+} (in the form of CaCl_2), where appropriate. Electrode resistance is set at 2-5 M Ω and electrodes are filled with the intracellular solution lacking mediator. Experiments are performed at room temperature from a holding potential of 0 mV. Voltage ramps (2.5 s) from -100 to 100 mV are acquired at a sampling frequency of 500 Hz. Current measured is proportional to the activity of TRICH in the assay.

Transport activity of TRICH is assayed by measuring uptake of labeled substrates into *Xenopus laevis* oocytes. Oocytes at stages V and VI are injected with TRICH mRNA (10 ng per oocyte) and incubated for 3 days at 18°C in OR2 medium (82.5mM NaCl, 2.5 mM KCl, 1mM CaCl_2 , 1mM MgCl_2 , 1mM Na_2HPO_4 , 5 mM Hepes, 3.8 mM NaOH, 50 $\mu\text{g/ml}$ gentamycin, pH 7.8) to allow expression of TRICH. Oocytes are then transferred to standard uptake medium (100mM NaCl, 2 mM KCl, 1mM CaCl_2 , 1mM MgCl_2 , 10 mM Hepes/Tris pH 7.5). Uptake of various substrates (e.g., amino acids, sugars, drugs, ions, and neurotransmitters) is initiated by adding labeled substrate (e.g. radiolabeled with ^3H , fluorescently labeled with rhodamine, etc.) to the oocytes. After incubating for 30 minutes, uptake is terminated by washing the oocytes three times in Na^+ -free medium, measuring the incorporated label, and comparing with controls. TRICH activity is proportional to the level of internalized labeled substrate.

ATPase activity associated with TRICH can be measured by hydrolysis of radiolabeled ATP- $[\gamma\text{-}^{32}\text{P}]$, separation of the hydrolysis products by chromatographic methods, and quantitation of the recovered ^{32}P using a scintillation counter. The reaction mixture contains ATP- $[\gamma\text{-}^{32}\text{P}]$ and varying amounts of TRICH in a suitable buffer incubated at 37°C for a suitable period of time. The reaction is terminated by acid precipitation with trichloroacetic acid and then neutralized with base, and an aliquot of the reaction mixture is subjected to membrane or filter paper-based chromatography to separate the reaction products. The amount of ^{32}P liberated is counted in a scintillation counter. The amount of radioactivity recovered is proportional to the ATPase activity of TRICH in the assay.

XVIII. Identification of TRICH Agonists and Antagonists

TRICH is expressed in a eukaryotic cell line such as CHO (Chinese Hamster Ovary) or HEK (Human Embryonic Kidney) 293. Ion channel activity of the transformed cells is measured in the presence and absence of candidate agonists or antagonists. Ion channel activity is assayed using patch clamp methods well known in the art or as described in Example XVII. Alternatively, ion channel activity is assayed using fluorescent techniques that measure ion flux across the cell membrane (Velicelebi, G. et al. (1999) *Meth. Enzymol.* 294:20-47; West, M.R. and C.R. Molloy (1996) *Anal. Biochem.* 241:51-58). These assays may be adapted for high-throughput screening using microplates. Changes in internal ion concentration are measured using fluorescent dyes such as the Ca^{2+} indicator Fluo-4 AM, sodium-sensitive dyes such as SBFI and sodium green, or the Cl^- indicator MQAE (all available from Molecular Probes) in combination with the FLIPR fluorimetric plate reading system (Molecular Devices). In a more generic version of this assay, changes in membrane potential caused by ionic flux across the plasma membrane are measured using oxonol dyes such as DiBAC₄ (Molecular Probes). DiBAC₄ equilibrates between the extracellular solution and cellular sites according to the cellular membrane potential. The dye's fluorescence intensity is 20-fold greater when bound to hydrophobic intracellular sites, allowing detection of DiBAC₄ entry into the cell (Gonzalez, J.E. and P.A. Negulescu (1998) *Curr. Opin. Biotechnol.* 9:624-631). Candidate agonists or antagonists may be selected from known ion channel agonists or antagonists, peptide libraries, or combinatorial chemical libraries.

Various modifications and variations of the described methods and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

Table 1

Project ID	IncYTE	Polypeptide SEQ ID NO:	Polypeptide ID	IncYTE	Polynucleotide SEQ ID NO:	Polynucleotide ID
1784775	1	1784775CD1	16	1784775CD1	16	1784775CB1
7473034	2	7473034CD1	17	7473034CD1	17	7473034CB1
1878581	3	1878581CD1	18	1878581CD1	18	1878581CB1
2246292	4	2246292CD1	19	2246292CD1	19	2246292CB1
5151730	5	5151730CD1	20	5151730CD1	20	5151730CB1
7472584	6	7472584CD1	21	7472584CD1	21	7472584CB1
7472536	7	7472536CD1	22	7472536CD1	22	7472536CB1
7473422	8	7473422CD1	23	7473422CD1	23	7473422CB1
2864715	9	2864715CD1	24	2864715CD1	24	2864715CB1
1734724	10	1734724CD1	25	1734724CD1	25	1734724CB1
1563237	11	1563237CD1	26	1563237CD1	26	1563237CB1
7473443	12	7473443CD1	27	7473443CD1	27	7473443CB1
7473438	13	7473438CD1	28	7473438CD1	28	7473438CB1
7474286	14	7474286CD1	29	7474286CD1	29	7474286CB1
7472589	15	7472589CD1	30	7472589CD1	30	7472589CB1

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability score	GenBank Homology
1	1784775CD1	94902730	4.00E-94	[Homo sapiens] multidrug resistance-associated protein 7
2	7473034CD1	92311724	9.00E-168	[Homo sapiens] aquaporin adipose Kuriyama H. et al. (1997) Biochem. Biophys. Res. Commun. 241, 53-58
3	1878581CD1	9545998	1.10E-164	[Rattus sp.] tricarboxylate carrier rats, liver, Peptide Mitochondrial partial, 357 aa Azzzi A. et al. (1993) J. Bioenerg. Biomembr. 25, 515-524
4	2246292CD1	96045150	0	[Rattus norvegicus] TAP-like ABC transporter Yamaguchi, Y. et al. (1999) FEBS Lett. 457, 231-236
5	5151730CD1	9475542	0	[Rattus norvegicus] glutamate receptor delta-1 subunit
		9220418	0	[Mus musculus] glutamate receptor channel subunit delta-1 Yamazaki M. et al. (1992) Biochem. Biophys. Res. Commun. 183:886-892
6	7472584CD1	97546839	0	[Cavia porcellus] potassium channel TASK1 Rajan, S. et al. (2000) J. Biol. Chem. 275:16650-16657
7	7472536CD1	92687858	4.30E-129	[Pseudopleuronectes americanus] renal organic anion transporter Wolff, N.A. et al. (1997) Expression cloning and characterization of a renal organic anion transporter from winter flounder, FEBS Lett 417:287-291
8	7473422CD1	99950945	0	[Pseudomonas aeruginosa] iron (III)-transport system permease HitB Stover, C.K. et al. (2000) Nature 406:959-964

Table 2 (cont.)

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability score	GenBank Homolog
9	2864715CD1	9531469	0	[Rattus norvegicus] renal osmotic stress-induced Na-Cl organic solute cotransporter
		93347922	5.00E-190	[Mus musculus] orphan transporter isoform AL2 Nash, S.R. et al. (1998) Cloning, gene structure and genomic localization of an orphan transporter from mouse kidney with six alternatively-spliced isoforms, Receptors Channels 6:113-128
10	1734724CD1	93980315	5.00E-46	[Oryctolagus cuniculus] hepatic sodium-dependent bile acid
11	1563237CD1	92208839	2.00E-125	[Rattus norvegicus] peptide/histidine transporter Yamashita, T. et al. (1997) Cloning and functional expression of a brain peptide/histidine transporter, J. Biol. Chem. 272:10205-10211
		97415511	0	[Homo sapiens] peptide transporter 3
12	7473443CD1	94186073	3.20E-280	[Mus musculus] calcium channel alpha-2-Delta-C subunit
13	7473438CD1	92465542	7.60E-84	[Homo sapiens] TWIK-related acid-sensitive K ⁺ channel Duprat, F. et al. (1997) TASK, a human background K ⁺ channel to sense external pH variations near physiological pH, EMBO J. 16:5464-5471
		g11228686	0	[Homo sapiens] two pore potassium channel RTP3.3

Table 2 (cont.)

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO:	Probability score	GenBank Homolog
14	7474286CD1	g306473	1.908-39	[Homo sapiens] calcium channel gamma subunit Powers, P.A. et al. (1993) Molecular characterization of the gene encoding the gamma subunit of the human skeletal muscle 1,4- dihydropyridine-sensitive Ca ²⁺ - channel (CACNLG), cDNA sequence, gene structure, and chromosomal location, J. Biol. Chem. 268:9275- 9279
15	7472589CD1	g11177514	0	[Homo sapiens] tandem pore domain potassium channel THIK-2 Rajan S. et al. (2001) J. Biol. Chem. 276:7302-7311

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1	1784775CD1	520	T343 S354 S404 T496 S345		Transmembrane domains: L25-L42; P120-Y142; L277-L297 L365-P392; Y465-M485 ABC transporter transmembrane region: Y242-E499 RESISTANCE; MULTIDRUG; SAURGLEISHMANIA; C3F10.11C; DM01742 P33527 195-653; P160-P201; Q221-V483 transmembrane domain: M46-L64 Major intrinsic protein (MIP): E31-Y276 MIP family proteins BL00221A:A43-V53 BL00221B:I92-T102 BL00221C:E179-D195 BL00221D:T225-I239 BL00221E:W258-G268 MAJOR INTRINSIC PROTEIN PR00783A:R39-S58 PR00783B:F78-T102 PR00783C:K115-I134 PR00783D:N178-Q196 PR00783E:G79-V101 PR00783F:W259-P279 AQUAPORIN ADIPOSE AQPAP TRANSPORT TRANSMEMBRANE PD078705:L277-F346 MIP FAMILY DM00228 F47862 15-263.K36-Y276 DM00228 I59266 15-263.K36-Y276 DM00228 F43549 340-587:R35-F279 DM00228 P11244 1-253:L42-Y276	HMW HMW-PPAM BLAST-DMO HMW HMW_PPAM BLINES_BLOCKS BLINES_PRINTS BLAST_PRODON BLAST_DOMO
2	7471034CD1	346	Y297 T11 S146 T192 T194 S252 S10	N65 N327		

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
3	1878591CD1	322	T30 T64 T143 T296 S309 S46 T134	N100 N124 N132 N135	PROTEIN TRANSMEMBRANE CHROMOSOME PUTATIVE TRANSPORTER P200698615-L251	BLAST_PRODOM
4	2246292CD1	723	T139 S161 T209 T311 T377 T469 S500 S720 S354 Y559 S28 S33 S46 T153 T181 S275 T367 S528 S552 S628 S659	N280 N465 N481 N556 N718	Abs transporter: L600-L614 ATP/GTP-binding site motif A: G496-S503 transmembrane domain V85-F104, V185-P204, L328-G347 ABC transporter: G489-G673 ABC transporter transmembrane region: L188-G450	MOTIFS HMMER HMMER_PPAM
					ABC transporters family BL00211A:L494-C505 BL00211B:L600-D631 ABC transporters family signature: I582-D631 ATP-BINDING TRANSPORT PROTEIN TRANSMEMBRANE GLYCOPROTEIN TRANSPORTER: PD000130:V229-A406	BLIMPS_BLOCKS PROFILESKAN BLAST_PRODOM
					WT2 PROTEIN DM07894 Q03519 1-685:L395-V675 DM07894 S25577 1-703:V422-M699 ABC TRANSPORTERS FAMILY DM00008 Q03518 502-717:V461-G673 DM00008 S13426 479-695:V461-G673	BLAST_DOMO

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
5	5151730CD1	1009	S17, S49, S57, S63, S104, S123, S171, S180, T211, T223, S263, S287, S355, T356, S364, T368, T400, S417, S424, T443, T438, S510, S528, S533, S659, T667, T681, T704, S724, S750, T752, S766, S785, S825, S857, T864, S882, S901, T939, Y220, Y679	N200, N422, N498	Metabotropic glutamate receptor: PR00593C; S518-S513 N-methyl-D-aspartate (NMDA) receptor signature: PR00177D; I632-S659 HYDROXYTRYPTAMINE 2A RECEPTOR (5-HT-2A)/(ISERCOTININ RECEPTOR) (5-HT-2): PR00177D; I632-S659 GLUTAMATE RECEPTOR SUBUNIT DELTA2 DELTA1 CHANNEL SUBTYPE DELTA: PD009864: M83-L437, PD009890:W854-I1009, PD000500:V565-V848, PD000273:L437-E550 GLUTAMATE RECEPTOR: DM08722 S28857 1-421:MI-N422, DM00247 S28857 627-897:S627-1898, DM08722 S28858 2-426:A20-N422, DM00247 S28858 628-900:S627-1898 Signal Peptide: MI-A20 Transmembrane Domain (transmembrane domain): L557-L583, S830-L847 Receptor family ligand binding region (AMF_Receptor): I12-S724	BLIMPS-PRINTS HMMER-SIGPEPT HMMER-TRANSMEMBRANE HMMER-PFAM

Table 3 (cont.)

SEQ ID NO.	Protein Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
6	7472584CD1	374	S31, S55, S127, S179, S251, S319, S334, T341, S360, S373,	N53	TASK K ⁺ channel signature: PR01095A: V6-V25, PR01095B: P26-K31, PR01095C: E63-Q77, PR01095D: Q126-K145, PR01095E: C146-T161, PR01095F: Q209-V230, PR01095G: G236-P254, PR01095H: R283-D306, PR01095K: P348-V374, CHANNEL IONIC TWIKRELATED ACIDSENSITIVE K ⁺ CTBKA F34D6.3 PROTEIN PUTATIVE POTASSIUM: PD013020: M1-N74 Signal Peptide (signal peptide): M1-V25 Transmembrane Domain (transmembrane domain): P225-V243 TASK K ⁺ channel (TWIK channel): M1-V374 Ion Transport Protein (ion trans): I54-L244	BLIPS-PRINTS
7	7472536CD1	589	S94, T121, T170, S192, S246, S308, S362, T376, T423, T513, S563, S564	N54, N74, N89, N96, N131, N149	Sugar transport proteins: DN00135 P38142 145-478: R230-E488 (p = 6.1e-06) Transmembrane domain: G400-M425, T424-M447 Sugar (and other) transporter: S133-A553	ELAST-PRODOM HMMER-SIGPEPT HMMER-TRANSMEMBRANE HMMER-PFAM ELAST-PRODOM HMMER-TRANSMEMBRANE HMMER-PFAM ELAST-PRODOM

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
8	7473422CD1	549	S3, S19, S49, S155, T182, T188, S189, S193 T449, T541		Binding-protein-dependent transport signature: BPD_transp: T182-E252	HMMER-PFAM
					Binding-protein-dependent transport systems inner membrane component: bpd_transp_inn_membr.prf: L432-E231	PROFILESSCAN
					Protein: iron iron III transport system, permease transport, transmembrane inner membrane HITB PD01934: K46-L183	BLAST-PRODOM
					DOMAIN: SFUB; IRON; TRANSPORT; PERIPLASMIC; DM05963 D64048 210-473: A316-L540	BLAST-DOMO
					BPD Transporter Integral Membrane protein signature: L183-P211, L443-P471	MOTIFS
					Transmembrane domain: P48-V70; N97-A117; M262-L279; Y315-S334; V356-V375	HMMER-TRANSMEMBRANE

Table 3 (cont.)

SEQ ID NO.	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
9	2864715CD1	634	S17, S30, S100, S143, S170, S186, T212, T219, T242, T267, T330, S483, T535, S561, S573	N158 N182 N258 N354 N368	SODIUM:NEUROTRANSMITTER SYMPORTER FAMILY: DM00572 S50998 19-616;S31-G614 Sodium:neurotransmitter: BL00610A: Q40-E89; BL00610B: W104-W153; BL00610C: W195-G246; BL00610D: E281-T313; BL00610E: A406-V448; BL00610F: M502-Q556; BL00610G: I576-F596 Sodium:neurotransmitter symporter family signature: neurotransm_transp_1: D36-F90 SODIUM/NEUROTRANSMITTER: PR00176A: Q40-L61; PR00176B: A69-L88; PR00176C: G113-Y139; PR00176D: A222-I239; PR00176E: V304-V324; PR00176F: M410-L429; PR00176G: S491-V511; PR00176H: Q531-V551 TRANSPORTER NEUROTRANSMITTER TRANSPORT TRANSMEMBRANE SYMPORT GLYCOPROTEIN SODIUM CHLORIDEDEPENDENT SODIUMDEPENDENT GABA: PD000448: C383-I606 Transmembrane domain: F70-L88, F123-F141, I305-V324, S417-V440, I462-Y480, T533-F550 Sodium neurotransmitter symporter family (SNF): R32-R332, L378-N608	BLAST-DOMO BLIMPS-BLOCKS PROFILESCAN BLIMPS-PRINTS BLAST-PRODOM HMMER- TRANSMEMBRANE HMMER-PFAM

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
10	I734724CD1	491	S227, S233, S459, T475	Ng, N18, N24, N180	SODIUM; ACID; BILE; TRANSPORTER DOMAIN; DMO1972 I38655 8-318:S251-Y439	BLAST-DOMO
					Transmembrane domain: V241-L261, L288-F307, W325-G343, M416-I435	HMNER
					Sodium Bile acid symporter family (SEF): L113-E345 (score=6.0, e-value=0.00017)	HMNER-PFAM
					ACID COTRANSPORTING POLYPEPTIDE; SODIUM/BILE COTRANSPORTER; NA+/BILE SODIUM/TAUROCHOLATE TRANSMEMBRANE TRANSPORT SYMPORTER: PD007533:L342-Y439	BLAST-PRODOM
					PTR2 family (putative protein transp.), proton/oligo: BL01022A:E42-L60; BL01022B:A72-L117; BL01022C:G164-V187; BL01022D:F199-V211; BL01022E:E416-S451	BLIMPS-BLOCKS
					Glucose transporter signature: PR00172C: V241-L261.	BLIMPS-PRINTS

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
11	1563237CD1	525	S281, S302, S367, T195, S174, T195, S281, S371	N61, N65, N178, N223, N356, N383	PEPTIDE OLIGOPEPTIDE TRANSPORTER; PROTEIN SYMPORT ISOFORM H/ PEPTIDE COTRANSPORTER; PD001550: Y101-S262	BLAST-PRODOM
					PTR2 FAMILY PROTON/OLIGOPEPTIDE SYMPORTERS: M01990 P46032 46-551:E42-L269	BLAST-DMO
					POT family (putative protein transporter) PTR2:Y101-S259; PTR2:N396-S447	HMER-PFAM
12	7473443CD1	1310	S52, T81, T95, S108, S109, S119, S176, S213, S330, T375, S410, S426, S459, S511, S550, S601, T725, S748, S761, T776, T780, S897, T918, S976, S1026, S1052, S1062, T1125, S1126, T1150, S1160, T1170, T1235, T1239, T1245, T1287	N16, N457, N661, N979, I124, I206	Similarity to calcium channels: DM06895 P54289 261-693:S759-L1187 CHANNEL CALCIUM PRECURSOR; IONIC TRANSMEMBRANE ION TRANSPORT; VOLTAGE-GATED DIHYDROPYRIDINESSENSITIVE LTYPE: PD013837:1631-K750	BLAST-DMO BLAST-PRODOM

Table 3 (cont.)

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
13	7473438CD1	400	S75 S101 S125 T162 S197 S249 S7 S105 S175 S332		Transmembrane domains: S75-F96; F295-V313 TASK K+ channel domain: M71-I400 TASK K+ channel signature PR01095; V76-V95; F96-K121; E133-Q147; Q196-C215; C216-V231; Q279-L300; G306-P324 CHANNEL PROTEIN IONIC POTASSIUM SUBUNIT K+ PUTATIVE SUBFAMILY X MEMBER PD021430; L229-V312	HMME HMME-PFAM BLIMPS-PRINTS
14	7474286CD1	260	T121 T133 T67 T88 T132 S195	N71 N113	Signal peptide: M1-A53 Transmembrane domains: V146-S164; F169-V187; V225-T242 CALCIUM CHANNEL GAMMA SUBUNIT DIHYDROPYRIDINESENSITIVE L1TYPE SKELETAL MUSCLE IONIC TRANSMEMBRANE PD016829; L31-C223 Transmembrane domains: F204-F223; P270-M293; F340-I359 TASK K+ channel domain: L32-E484 (Score = -115.2; E-value = 0.0078)	SPSCAN HMME BLAST-PRODOM HMME-PFAM
15	7472589CD1	489	S257 S373 S424 S58 S62 S260 T295 S416 S429 S2 S5 S13 T36 T75 T149 T156 T252 S436	N78		

Table 4

Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
16	1784775CB1	1735	233-901, 1- 59, 1252- 1513	60118407D2 60209102U1 67560533J1 (SINTFER02) 64318533H1 (LUNGNON07) GBI:96850429_000023. fasta.edit.comp	948 1053 1057 324 1	1303 1416 1735 1021 771
17	7473034CB1	1041	158-280	GBI:96742097_000003. fasta.edit	1	1041
18	1878581CB1	2367	1-42, 1275- 2367	71456748V1 71461790V1 71465463V1 71464816V1 71465977V1 71456476V1 71466268V1 4202150H1 (BRAITUT29) 5531403F6 (HEARF05)	228 1602 300 1499 807 1723 908 1 1747	852 2235 863 2193 1533 2367 1564 280 2162
19	2246292CB1	3343	1-867, 2606-2809, 1501-1983, 960-1013	60111369D2 65396013 3811252P6 (CONFUT01) 2246292R6 (HIFONON02) 6890917J1 (BRAITDR03) 7168417H1 (MCLRNOC01) 4691312T6 (BRAENOT02) 6120730H1 (BRAHNON05) 3334167H1 (BRAIFET01) 4691312P6 (BRAENOT02) 5014991H1 (BRAXNOT03) 2395385F6 (THELAZT01)	1 2884 1387 2978 650 100 2689 2365 549 1959 1713 1026	473 3343 1948 3333 1293 648 3320 3044 818 2552 1968 1490

Table 4 (cont.)

Polynucleotide Seq ID NO:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
20	5151730CB1	3517	863-1689, 1-113, 3288-3339	6885903JL (BHAHTRD03) 5151730F6 (HEARFET03) 1675978T6 (BLADNOT05) 3907290H1 (LUNGNOT23) 4406394H1 (PROSDIT01) 6974079F8 (BHAHTRD04) 7288330H1 (BPAIFER06) 6250428F8 (LUNPNU02) 7326591H2 (SPINUT02) 7174089H1 (BRSTTMC01) 3486391G (BFIGNOT01) 6851642H1 (BPAITDR03) 445149H1 (TUPPNOT02) GBL.66164987.Faw 92001843 7693385H2 (LNUDTUE01) 7693385J2 (LNUDTUE01) GNN.55776606_000002_002 .edit GNN.95525050_000002_004 70870289V1 70529541V1 6783918H1 (SINITMC01) 6829267J1 (SINTHOR01) 70792490V1 6783311H1 (SINITMC01) 6783433H2 (SINITMC01) 1734724F6 (COLANNOT22) 60123116D1 4222214H1 (PANCNOT07) 3455639T7 (ADRENU07) 3495639F6 (ADRENU07) GNN.67344269_000021_002 6147666H1 (BRANDIT03)	2374 604 2973 3381 2229 1 1855 183 804 1092 2821 1458 959 1 946 1 253 1 1 1771 1 550 1254 1088 2213 1710 1 1063 902 1942 1525 963 69 1	2952 927 3502 3517 2485 582 2316 856 1409 1683 3440 1993 1128 1122 1248 743 879 1770 2544 2335 1154 1738 1616 2871 2261 466 1704 987 2141 2120 1551 1544 545
21	7472584CB1	1248	751-945, 96-219			
22	7472536CB1	1770	1-481, 1136-1770, 573-1030			
23	7473422CB1	2544	1-2544			
24	2864715CB1	2871	1-32, 750- 1421, 1950- 2871			
25	1734724CB1	2141	658-817			

Table 4 (cont.)

Polynucleotide Seq ID No:	Incyte Polynucleotide ID	Sequence Length	Selected Fragment(s)	Sequence Fragments	5' Position	3' Position
26	1563237CB1	1902	1-1283	7127258H1 (COLNDIV01) 2207658F6 (SINTFE03) 1563237T6 (SPENOT04) 2118017T6 (BRSTUT02) 7679862H1 (BRAFTUE01)	589 818 1332 1410 1	1258 1379 1899 1902 722
27	7473443CB1	4125	1-41, 3763- 4125, 407- 435, 491- 2446	71153376V1 FL722155_00001.raw 71395953V1 GB1.G3810573.raw 71151716V1 71158575V1	3186 1864 3713 1 3338 2741	3857 2161 4125 804 3892 3346
28	7473438CB1	2460	1628-1717, 1-374, 988- 1020, 515- 540, 2181- 2460, 1182- 1278	GB1.G3810573.edit 6758949H1 (HEAONOR01) 2230573F6 (PROSNOT16) FL767399_00001 Cpg_991027.B15_masked_ fa.Contig52563	634 1907 1541 722 1	3099 2460 2120 1974 1182
29	7474286CB1	896	86-487	GNN.G7523629_000001_ 002	1	896
30	7472589CB1	2080	1-492, 1002-1177, 618-691	7289171H1 (BRAIFER06) GNN.G6560849_000001_ 002	1 285	599 2080

Table 5

Polynucleotide Seq ID NO:	Incyte Project ID	Representative Library
16	1784775CB1	BRAINOT10
18	1878581CB1	SELPFET02
19	2246292CB1	HIPONON02
20	5151730CB1	PROSTW07
21	7472584CB1	TYMNOT02
22	7472536CB1	LANOTUE01
24	2864715CB1	SINITUT03
25	1734724CB1	ADRETUT07
26	1563237CB1	DRGLNOT01
27	7473443CB1	THPLAZS08
28	7473438CB1	PROSNOT16
29	7474286CB1	PHYNOT03
30	7472589CB1	BRAIFER06

Table 6

Library	Vector	Library Description
ADREUT07	pINCY	Library was constructed using RNA isolated from adrenal tumor tissue removed from a 43-year-old Caucasian female during a unilateral adrenalectomy. Pathology indicated pheochromocytoma.
BRAIFER06	PCDNA2.1	This random primed library was constructed using RNA isolated from brain tissue removed from a Caucasian male fetus who was stillborn with a hypoplastic left heart at 23 weeks' gestation. Serologies were negative.
BRAINOT10	pINCY	Library was constructed using RNA isolated from diseased cerebellum tissue removed from the brain of a 74-year-old Caucasian male, who died from Alzheimer's disease.
DRGLN0701	pINCY	Library was constructed using RNA isolated from dorsal root ganglion tissue removed from the cervical spine of a 32-year-old Caucasian male who died from acute pulmonary edema and bronchopneumonia, bilateral pleural and pericardial effusions, and malignant lymphoma (natural killer cell type). Patient history included probable cytomegalovirus, infection, hepatic congestion and steatosis, splenomegaly, hemorrhagic cystitis, thyroid hemorrhage, and Bell's palsy. Surgeries included colonoscopy, large intestine biopsy, adenotonsillectomy, and nasopharyngeal endoscopy and biopsy; treatment included radiation therapy.
HIFON0802	PSPORT1	This normalized hippocampus library was constructed from 1.13M independent clones from a hippocampus tissue library. RNA was isolated from the hippocampus tissue of a 72-year-old Caucasian female who died from an intracranial bleed. Patient history included nose cancer, hypertension, and arthritis. The normalization and hybridization conditions were adapted from Soares et al. (PNAS (1994) 91:9228).
LANODTUE01	PCDNA2.1	This 5' biased random primed library was constructed using RNA isolated from lymph node tumor tissue removed from a 67-year-old Caucasian male during regional lymph node excision, open biopsy of tongue, and partial glossectomy. Pathology indicated metastatic basaloid squamous cell carcinoma. Pathology for the associated tumor tissue indicated basaloid squamous cell carcinoma forming an ulcerated mass in the base of the tongue. Tumor deeply invaded the underlying skeletal muscle. The patient presented with tongue cancer. Patient history included benign hypertension and alcohol abuse. The patient was not taking any medications. Family history included diabetes in the son.

Table 6 (cont.)

Library	Vector	Library Description
PROSNOT16	pINCY	Library was constructed using RNA isolated from diseased prostate tissue removed from a 68-year-old Caucasian male during a radical prostatectomy. Pathology indicated adenofibromatous hyperplasia. Pathology for the associated tumor tissue indicated an adenocarcinoma (Gleason grade 3+4). The patient presented with elevated prostate specific antigen (PSA). During this hospitalization, the patient was diagnosed with myasthenia gravis. Patient history included osteoarthritis, and type II diabetes. Family history included benign hypertension, acute myocardial infarction, hyperlipidemia, and arteriosclerotic coronary artery disease.
PROSNTW07	pINCY	The library was constructed using RNA isolated from diseased prostate tissue removed from a 73-year-old Caucasian male during radical prostatectomy, closed prostatic biopsy, and regional lymph node excision. Pathology indicated adenofibromatous hyperplasia. Pathology for the associated tumor tissue indicated adenocarcinoma, Gleason 3+3, involving the left side peripherally and anteriorly. The tumor perforated the capsule to involve periprostatic tissue and anterior surgical margin on the left. The patient presented with elevated prostate-specific antigen. Patient history included bladder cancer, speech disturbance and acquired spondylolisthesis. Family history included benign hypertension and cerebrovascular disease.
PTHYNG003	pINCY	Library was constructed using RNA isolated from the left parathyroid tissue of a 69-year-old Caucasian female during a partial parathyroidectomy. Pathology indicated hyperplasia. The patient presented with primary hyperparathyroidism.
SINTUT003	pINCY	Library was constructed using RNA isolated from ileal tumor tissue obtained from a 49-year-old Caucasian female during destruction of peritoneal tissue, peritoneal adhesiolysis, ileum resection, and permanent colectomy. Pathology indicated grade 4 adenocarcinoma. Patient history included benign hypertension. Previous surgeries included total abdominal hysterectomy, bilateral salpingo-oophorectomy, regional lymph node excision, an incidental appendectomy, and dilation and curettage. Family history included benign hypertension, and cerebrovascular disease. Hyperlipidemia, arteriosclerotic coronary artery disease, hyperlipidemia, type II diabetes, and stomach cancer.
SPINFT002	pINCY	Library was constructed using RNA isolated from spleen tissue removed from a Caucasian male fetus, who died at 23 weeks' gestation.

Table 6 (cont.)

Library	Vector	Library Description
THPIAZS08	PSFORT1	<p>This subtracted THP-1 promonocyte cell line library was constructed using 5.76 x 1e6 clones from a 5-aza-2'-deoxycytidine (AZ) treated THP-1 cell library. Starting RNA was made from THP-1 promonocyte cells treated for three days with 0.8 micromolar AZ. The donor had acute monocytic leukemia. The hybridization probe for subtraction was derived from a similarly constructed library, made from 1 microgram of polyA RNA isolated from untreated THP-1 cells. 5.76 million clones from the AZ-treated THP-1 cell library were then subjected to two rounds of subtractive hybridization with 5 million clones from the untreated THP-1 cell library. Subtractive hybridization conditions were based on the methodologies of Swaroop et al., NAR (1991) 19:1954, and Bonaldo et al., Genome Research (1996) 6:791.</p>
TLYMN0702	Lambda Unizap	<p>Library was constructed using RNA isolated from non-adherent peripheral blood mononuclear cells. The blood was obtained from unrelated male and female donors and treated with LPS for 0 hours. Cells from each donor were purified on Ficol1 Hypaque, then harvested by centrifugation, lysed in a buffer containing GUSCN, and spun through CsCl to obtain RNA for library construction. PolyA RNA was isolated using a Qiagen Oligotex kit. cDNA synthesis was initiated using an XhoI-oligo(dT) primer. Double-stranded cDNA was bluntended, ligated to EcoRI adaptors, digested with XhoI, size-selected, and cloned into the XhoI and EcoRI sites of the Lambda Unizap vector.</p>

Table 7

Program	Description	Reference	Parameter Threshold
ABI/FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.	ESTs: Probability value= 1.0E-8 or less Full Length sequences: Probability value= 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises at least five functions: fasta, tfasta, fastx, ifastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad. Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	ESTs: fasta E value=1.0E-6 Assembled ESTs: fasta identity= 95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less Full Length sequences: fastx score=100 or greater
BLIMPS	A Blocks IMPROVED Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Atwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	Probability value= 1.0E-3 or less
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM.	Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322; Durbin, R. et al. (1998) Our World View, in a Nutshell, Cambridge Univ. Press, pp. 1-350.	PFAM hits: Probability value= 1.0E-3 or less Signal peptide hits: Score= 0 or greater

Table 7 (cont.)

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gritskov, M. et al. (1988) CABIOS 4:61-66; Gritskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality score \geq GCG-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, R. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score = 120 or greater; Match length = 36 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score = 3.5 or greater
TMAP	A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.	Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:363-371.	
TMHMMER	A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.	Sonnhammer, E.L. et al. (1998) Proc. Sixth Intl. Conf. on Intelligent Systems for Mol. Biol., Glasgow et al., eds. The Am. Assoc. for Artificial Intelligence Press, Menlo Park, CA, pp. 175-182.	
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

What is claimed is:

1. An isolated polypeptide selected from the group consisting of:
 - a) a polypeptide comprising an amino acid sequence selected from the group consisting of
5 SEQ ID NO:1-15,
 - b) a naturally occurring polypeptide comprising an amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:4 and SEQ ID NO:6-15,
 - c) a naturally occurring polypeptide comprising an amino acid sequence at least 92%
10 identical to the amino acid sequence of SEQ ID NO:2,
 - d) a naturally occurring polypeptide comprising an amino acid sequence at least 96% identical to the amino acid sequence of SEQ ID NO:3,
 - e) a naturally occurring polypeptide comprising an amino acid sequence at least 99% identical to the amino acid sequence of SEQ ID NO:5,
 - 15 f) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, and
 - g) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15.
- 20 2. An isolated polypeptide of claim 1 selected from the group consisting of SEQ ID NO:1-15.
3. An isolated polynucleotide encoding a polypeptide of claim 1.
- 25 4. An isolated polynucleotide encoding a polypeptide of claim 2.
5. An isolated polynucleotide of claim 4 selected from the group consisting of SEQ ID NO:16-30.
- 30 6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.
7. A cell transformed with a recombinant polynucleotide of claim 6.

8. A transgenic organism comprising a recombinant polynucleotide of claim 6.
9. A method for producing a polypeptide of claim 1, the method comprising:
- a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said
 - 5 cell is transformed with a recombinant polynucleotide, and said recombinant polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and
 - b) recovering the polypeptide so expressed.
10. An isolated antibody which specifically binds to a polypeptide of claim 1.
11. An isolated polynucleotide selected from the group consisting of:
- a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting
 - of SEQ ID NO:16-30,
 - 15 b) a naturally occurring polynucleotide comprising a polynucleotide sequence at least 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:16-30,
 - c) a polynucleotide complementary to the polynucleotide of a),
 - d) a polynucleotide complementary to the polynucleotide of b), and
 - 20 e) an RNA equivalent of a)-d).
12. An isolated polynucleotide comprising at least 60 contiguous nucleotides of a polynucleotide of claim 11.
13. A method for detecting a target polynucleotide in a sample, said target polynucleotide
- 25 having a sequence of a polynucleotide of claim 11, the method comprising:
 - a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and
 - 30 b) detecting the presence or absence of said hybridization complex, and, optionally, if present, the amount thereof.
14. A method of claim 13, wherein the probe comprises at least 60 contiguous nucleotides.

15. A method for detecting a target polynucleotide in a sample, said target polynucleotide having a sequence of a polynucleotide of claim 11, the method comprising:

a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and

5 b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof, and, optionally, if present, the amount thereof.

16. A composition comprising a polypeptide of claim 1 and a pharmaceutically acceptable excipient.

10

17. A composition of claim 16, wherein the polypeptide has an amino acid sequence selected from the group consisting of SEQ ID NO:1-15.

18. A method for treating a disease or condition associated with decreased expression of functional TRICH, comprising administering to a patient in need of such treatment the composition of claim 16.

19. A method for screening a compound for effectiveness as an agonist of a polypeptide of claim 1, the method comprising:

20 a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
b) detecting agonist activity in the sample.

20. A composition comprising an agonist compound identified by a method of claim 19 and a pharmaceutically acceptable excipient.

25

21. A method for treating a disease or condition associated with decreased expression of functional TRICH, comprising administering to a patient in need of such treatment a composition of claim 20.

22. A method for screening a compound for effectiveness as an antagonist of a polypeptide of claim 1, the method comprising:

30 a) exposing a sample comprising a polypeptide of claim 1 to a compound, and
b) detecting antagonist activity in the sample.

23. A composition comprising an antagonist compound identified by a method of claim 22 and a pharmaceutically acceptable excipient.

24. A method for treating a disease or condition associated with overexpression of functional TRICH, comprising administering to a patient in need of such treatment a composition of claim 23.

25. A method of screening for a compound that specifically binds to the polypeptide of claim 1, said method comprising the steps of:

- a) combining the polypeptide of claim 1 with at least one test compound under suitable conditions, and
- b) detecting binding of the polypeptide of claim 1 to the test compound, thereby identifying a compound that specifically binds to the polypeptide of claim 1.

26. A method of screening for a compound that modulates the activity of the polypeptide of claim 1, said method comprising:

- a) combining the polypeptide of claim 1 with at least one test compound under conditions permissive for the activity of the polypeptide of claim 1,
- b) assessing the activity of the polypeptide of claim 1 in the presence of the test compound, and
- c) comparing the activity of the polypeptide of claim 1 in the presence of the test compound with the activity of the polypeptide of claim 1 in the absence of the test compound, wherein a change in the activity of the polypeptide of claim 1 in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide of claim 1.

27. A method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a sequence of claim 5, the method comprising:

- a) exposing a sample comprising the target polynucleotide to a compound, under conditions suitable for the expression of the target polynucleotide,
- b) detecting altered expression of the target polynucleotide, and
- c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

28. A method for assessing toxicity of a test compound, said method comprising:

- a) treating a biological sample containing nucleic acids with the test compound;

- b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide of claim 11 under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide comprising a polynucleotide sequence of a polynucleotide of claim 11 or fragment thereof;
- c) quantifying the amount of hybridization complex; and
- d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

29. A diagnostic test for a condition or disease associated with the expression of TRICH in a biological sample comprising the steps of:

- a) combining the biological sample with an antibody of claim 10, under conditions suitable for the antibody to bind the polypeptide and form an antibody:polypeptide complex; and
- b) detecting the complex, wherein the presence of the complex correlates with the presence of the polypeptide in the biological sample.

30. The antibody of claim 10, wherein the antibody is:

- a) a chimeric antibody,
- b) a single chain antibody,
- c) a Fab fragment,
- d) a F(ab')₂ fragment, or
- e) a humanized antibody.

31. A composition comprising an antibody of claim 10 and an acceptable excipient.

32. A method of diagnosing a condition or disease associated with the expression of TRICH in a subject, comprising administering to said subject an effective amount of the composition of claim

33. A composition of claim 31, wherein the antibody is labeled.

34. A method of diagnosing a condition or disease associated with the expression of TRICH

in a subject, comprising administering to said subject an effective amount of the composition of claim 33.

35. A method of preparing a polyclonal antibody with the specificity of the antibody of claim 10 comprising:

- a) immunizing an animal with a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, or an immunogenic fragment thereof, under conditions to elicit an antibody response;
- b) isolating antibodies from said animal; and
- c) screening the isolated antibodies with the polypeptide, thereby identifying a polyclonal antibody which binds specifically to a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15.

36. An antibody produced by a method of claim 35.

37. A composition comprising the antibody of claim 36 and a suitable carrier.

38. A method of making a monoclonal antibody with the specificity of the antibody of claim 10 comprising:

- a) immunizing an animal with a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15, or an immunogenic fragment thereof, under conditions to elicit an antibody response;
- b) isolating antibody producing cells from the animal;
- c) fusing the antibody producing cells with immortalized cells to form monoclonal antibody-producing hybridoma cells;
- d) culturing the hybridoma cells; and
- e) isolating from the culture monoclonal antibody which binds specifically to a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15.

39. A monoclonal antibody produced by a method of claim 38.

40. A composition comprising the antibody of claim 39 and a suitable carrier.

41. The antibody of claim 10, wherein the antibody is produced by screening a Fab

expression library.

42. The antibody of claim 10, wherein the antibody is produced by screening a recombinant immunoglobulin library.

5

43. A method for detecting a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15 in a sample, comprising the steps of:

- a) incubating the antibody of claim 10 with a sample under conditions to allow specific binding of the antibody and the polypeptide; and
- 10 b) detecting specific binding, wherein specific binding indicates the presence of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15 in the sample.

44. A method of purifying a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15 from a sample, the method comprising:

- a) incubating the antibody of claim 10 with a sample under conditions to allow specific binding of the antibody and the polypeptide; and
- b) separating the antibody from the sample and obtaining the purified polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-15.

20

45. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:1.

46. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:2.

25 47. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:3.

48. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:4.

49. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:5.

30

50. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:6.

51. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:7.

52. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:8.
53. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:9.
- 5 54. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:10.
55. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:11.
56. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:12.
- 10 57. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:13.
58. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:14.
- 15 59. A polypeptide of claim 1, comprising the amino acid sequence of SEQ ID NO:15.
60. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:16.
- 20 61. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:17.
62. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:18.
- 25 63. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:19.
64. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:20.
- 30 65. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID NO:21.

66. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID
NO:22.

67. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID
5 NO:23.

68. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID
NO:24.

10 69. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID
NO:25.

70. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID
NO:26.

15 71. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID
NO:27.

72. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID
20 NO:28.

73. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID
NO:29.

25 74. A polynucleotide of claim 11, comprising the polynucleotide sequence of SEQ ID
NO:30.

<110> INCYTE GENOMICS, INC.
 REDDY, Roopa
 THORNTON, Michael
 BOROWSKY, Mark L.
 TANG, Y. Tom
 KHAN, Farrah A.
 TRIBOULEY, Catherine M.
 GANDHI, Ameena R.
 YAO, Monique G.
 SANTANWALA, Madhu S.
 BAUGHN, Mariah R.
 NGUYEN, Dannie B.
 POLICKY, Jennifer L.
 YUE, Henry
 SEILHAMER, Jeffrey J.
 WALIA, Narinder K.
 LAL, Preeti
 KEARNEY, Liam
 WALSH, Roderick T.
 LU, Dyung Aina M.
 LU, Yan
 GREENE, Barrie D.
 RAUMANN, Brigitte E.
 PATTERSON, Chandra

<120> TRANSPORTERS AND ION CHANNELS

<130> PI-0067 PCT

<140> To Be Assigned

<141> Herewith

<150> 60/195,595; 60/196,872; 60/199,020; 60/200,552; 60/202,348;
 60/203,495

<151> 2000-04-06; 2000-04-12; 2000-04-20; 2000-04-28; 2000-05-05;
 2000-05-11

<160> 30

<170> PERL Program

<210> 1

<211> 520

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 1784775CD1

<400> 1

Met	Cys	Leu	Leu	Val	Phe	Pro	Leu	Val	Pro	Arg	Ser	Pro	Asp	Tyr
1				5					10					15
Ile	Leu	Pro	Cys	Ser		Pro	Gly	Trp	Arg	Leu	Arg	Leu	Ala	Ser
				20					25					30
Phe	Leu	Leu	Ser	Val	Phe	Pro	Leu	Leu	Asp	Leu	Leu	Pro	Val	Ala
				35					40					45
Leu	Pro	Pro	Gly	Ala	Gly	Pro	Gly	Pro	Ile	Gly	Leu	Glu	Val	Leu
				50					55					60
Ala	Gly	Cys	Val	Ala	Ala	Val	Ala	Trp	Ile	Ser	His	Ser	Leu	Ala
				65					70					75
Leu	Trp	Val	Leu	Ala	His	Ser	Pro	His	Gly	His	Ser	Arg	Gly	Pro
				80					85					90

Leu	Ala	Leu	Ala	Leu	Val	Ala	Leu	Leu	Pro	Ala	Pro	Ala	Leu	Val
				95					100					105
Leu	Thr	Val	Leu	Trp	His	Cys	Gln	Arg	Gly	Thr	Leu	Leu	Pro	Pro
				110					115					120
Leu	Leu	Pro	Gly	Pro	Met	Ala	Arg	Leu	Cys	Leu	Leu	Ile	Leu	Gln
				125					130					135
Leu	Ala	Ala	Leu	Leu	Ala	Tyr	Ala	Leu	Gly	Trp	Ala	Ala	Pro	Gly
				140					145					150
Gly	Pro	Arg	Glu	Pro	Trp	Ala	Gln	Glu	Pro	Leu	Leu	Pro	Glu	Asp
				155					160					165
Gln	Glu	Pro	Glu	Val	Ala	Glu	Asp	Gly	Glu	Ser	Trp	Leu	Ser	Arg
				170					175					180
Phe	Ser	Tyr	Ala	Trp	Leu	Ala	Pro	Leu	Leu	Ala	Arg	Gly	Ala	Cys
				185					190					195
Gly	Glu	Leu	Arg	Gln	Pro	Gln	Asp	Ile	Cys	Arg	Leu	Pro	His	Arg
				200					205					210
Leu	Gln	Pro	Thr	Tyr	Leu	Ala	Arg	Val	Phe	Gln	Ala	His	Trp	Gln
				215					220					225
Glu	Gly	Ala	Arg	Leu	Trp	Arg	Ala	Leu	Tyr	Gly	Ala	Phe	Gly	Arg
				230					235					240
Cys	Tyr	Leu	Ala	Leu	Gly	Leu	Leu	Lys	Leu	Val	Gly	Thr	Met	Leu
				245					250					255
Gly	Phe	Ser	Gly	Pro	Leu	Leu	Leu	Ser	Leu	Leu	Val	Gly	Phe	Leu
				260					265					270
Glu	Glu	Gly	Gln	Glu	Pro	Leu	Ser	His	Gly	Leu	Leu	Tyr	Ala	Leu
				275					280					285
Gly	Leu	Ala	Gly	Gly	Ala	Val	Leu	Gly	Ala	Val	Leu	Gln	Asn	Gln
				290					295					300
Tyr	Gly	Tyr	Glu	Val	Tyr	Lys	Val	Thr	Leu	Gln	Ala	Arg	Gly	Ala
				305					310					315
Val	Leu	Asn	Ile	Leu	Tyr	Cys	Lys	Ala	Leu	Gln	Leu	Gly	Pro	Ser
				320					325					330
Arg	Pro	Pro	Thr	Gly	Glu	Ala	Leu	Asn	Leu	Leu	Gly	Thr	Asp	Ser
				335					340					345
Glu	Arg	Leu	Leu	Asn	Phe	Ala	Gly	Ser	Phe	His	Glu	Ala	Trp	Gly
				350					355					360
Leu	Pro	Leu	Gln	Leu	Ala	Ile	Thr	Leu	Tyr	Leu	Leu	Tyr	Gln	Gln
				365					370					375
Val	Gly	Val	Ala	Phe	Val	Gly	Gly	Leu	Ile	Leu	Ala	Leu	Leu	Leu
				380					385					390
Val	Pro	Val	Asn	Lys	Val	Ile	Ala	Thr	Arg	Ile	Met	Ala	Ser	Asn
				395					400					405
Gln	Glu	Met	Leu	Gln	His	Lys	Asp	Ala	Arg	Val	Lys	Leu	Val	Thr
				410					415					420
Glu	Leu	Leu	Ser	Gly	Ile	Arg	Val	Ile	Lys	Phe	Cys	Gly	Trp	Glu
				425					430					435
Gln	Ala	Leu	Gly	Ala	Arg	Val	Glu	Ala	Cys	Arg	Ala	Arg	Glu	Leu
				440					445					450
Gly	Arg	Leu	Arg	Val	Ile	Lys	Tyr	Leu	Asp	Ala	Ala	Cys	Val	Tyr
				455					460					465
Leu	Trp	Ala	Ala	Leu	Pro	Val	Val	Ile	Ser	Ile	Val	Ile	Phe	Ile
				470					475					480
Thr	Tyr	Val	Leu	Met	Gly	His	Gln	Leu	Thr	Ala	Thr	Lys	Val	Arg
				485					490					495
Thr	Arg	Lys	Glu	Gly	Asp	Gln	His	Gln	Gly	Asp	Phe	Ser	Glu	Val
				500					505					510
Lys	Thr	Glu	Ala	Trp	Ala	Leu	Ser	Ala	Gly					
				515					520					

<210> 2

<211> 346

<212> PRT

<213> Homo sapiens

<220>
 <231> misc_feature
 <223> Incyte ID No: 7473034CD1

<400> 2
 Met Val Gln Ala Ser Gly His Arg Arg Ser Thr Arg Gly Ser Lys
 1 5 10 15
 Met Val Ser Trp Ser Val Ile Ala Lys Ile Gln Glu Ile Trp Cys
 20 25 30
 Glu Glu Asp Glu Arg Lys Met Val Arg Glu Phe Leu Ala Glu Phe
 35 40 45
 Met Ser Thr Tyr Val Met Met Val Phe Gly Leu Gly Ser Val Ala
 50 55 60
 His Met Val Leu Asn Lys Thr Tyr Gly Ser Tyr Leu Gly Val Asn
 65 70 75
 Leu Gly Phe Gly Phe Gly Val Thr Met Gly Val His Val Ala Gly
 80 85 90
 Arg Ile Ser Gly Ala His Met Asn Ala Ala Val Thr Phe Thr Asn
 95 100 105
 Cys Ala Leu Gly Arg Val Pro Trp Arg Lys Phe Pro Val His Val
 110 115 120
 Leu Gly Gln Phe Leu Gly Ser Phe Leu Ala Ala Thr Ile Tyr
 125 130 135
 Ser Leu Phe Tyr Ser Ala Ile Leu His Phe Ser Gly Gly Glu Leu
 140 145 150
 Met Val Thr Gly Pro Phe Ala Thr Ala Gly Ile Phe Ala Thr Tyr
 155 160 165
 Leu Pro Asp His Met Thr Leu Trp Arg Gly Phe Leu Asn Glu Glu
 170 175 180
 Trp Leu Thr Arg Met Leu Gln Leu Cys Leu Phe Thr Ile Thr Asp
 185 190 195
 Gln Glu Asn Asn Pro Ala Leu Pro Gly Thr His Ala Leu Val Ile
 200 205 210
 Ser Ile Leu Val Val Ile Ile Arg Val Ser His Gly Ile Asn Thr
 215 220 225
 Gly Tyr Ala Ile Asn Pro Ser Arg Asp Pro Pro Pro Ser Ile Phe
 230 235 240
 Thr Phe Ile Ala Gly Trp Gly Lys Gln Val Phe Ser Asp Gly Glu
 245 250 255
 Asn Trp Trp Trp Val Pro Val Val Ala Pro Leu Leu Gly Ala Ser
 260 265 270
 Leu Gly Gly Ile Ile Tyr Leu Val Phe Ile Gly Ser Thr Ile Pro
 275 280 285
 Arg Glu Pro Leu Lys Leu Glu Asp Ser Val Ala Tyr Glu Asp His
 290 295 300
 Gly Ile Thr Val Leu Pro Lys Met Gly Ser His Glu Pro Met Ile
 305 310 315
 Ser Pro Leu Thr Leu Ile Ser Val Ser Leu Ala Asn Arg Ser Ser
 320 325 330
 Val His Ser Ala Pro Pro Leu His Glu Ser Met Ala Leu Glu His
 335 340 345
 Phe

<210> 3
 <211> 322
 <312> PRT
 <313> Homo sapiens

<220>
 <231> misc_feature
 <223> Incyte ID No: 1878581CD1

```

<400> 3
Met Ser Gly Glu Leu Pro Pro Asn Ile Asn Ile Lys Glu Pro Arg
1      5      10      15
Trp Asp Gln Ser Thr Phe Ile Gly Arg Ala Asn His Phe Phe Thr
20      25      30
Val Thr Asp Pro Arg Asn Ile Leu Leu Thr Asn Glu Gln Leu Glu
35      40      45
Ser Ala Arg Lys Ile Val His Asp Tyr Arg Gln Gly Ile Val Pro
50      55      60
Pro Gly Leu Thr Glu Asn Glu Leu Trp Arg Ala Lys Tyr Ile Tyr
65      70      75
Asp Ser Ala Phe His Pro Asp Thr Gly Glu Lys Met Ile Leu Ile
80      85      90
Gly Arg Met Ser Ala Gln Val Pro Met Asn Met Thr Ile Thr Gly
95      100     105
Cys Met Met Thr Phe Tyr Arg Thr Thr Pro Ala Val Leu Phe Trp
110     115     120
Gln Trp Ile Asn Gln Ser Phe Asn Ala Val Val Asn Tyr Thr Asn
125     130     135
Arg Ser Gly Asp Ala Pro Leu Thr Val Asn Glu Leu Gly Thr Ala
140     145     150
Tyr Val Ser Ala Thr Thr Gly Ala Val Ala Thr Ala Leu Gly Leu
155     160     165
Asn Ala Leu Thr Lys His Val Ser Pro Leu Ile Gly Arg Phe Val
170     175     180
Pro Phe Ala Ala Val Ala Ala Ala Asn Cys Ile Asn Ile Pro Leu
185     190     195
Met Arg Gln Arg Glu Leu Lys Val Gly Ile Pro Val Thr Asp Glu
200     205     210
Asn Gly Asn Arg Leu Gly Glu Ser Ala Asn Ala Ala Lys Gln Ala
215     220     225
Ile Thr Gln Val Val Ser Arg Ile Leu Met Ala Ala Pro Gly
230     235     240
Met Ala Ile Pro Pro Phe Ile Met Asn Thr Leu Glu Lys Lys Ala
245     250     255
Phe Leu Lys Arg Phe Pro Trp Met Ser Ala Pro Ile Gln Val Gly
260     265     270
Leu Val Gly Phe Cys Leu Val Phe Ala Thr Pro Leu Cys Cys Ala
275     280     285
Leu Phe Pro Gln Lys Ser Ser Met Ser Val Thr Ser Leu Glu Ala
290     295     300
Glu Leu Gln Ala Lys Ile Gln Glu Ser His Pro Glu Leu Arg Arg
305     310     315
Val Tyr Phe Asn Lys Gly Leu
320

```

```

<210> 4
<211> 723
<212> PRT
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<223> Incyte ID No: 2246292CD1

```

```

<400> 4
Met Arg Leu Trp Lys Ala Val Val Val Thr Leu Ala Phe Met Ser
1      5      10      15
Val Asp Ile Cys Val Thr Thr Ala Ile Tyr Val Phe Ser His Leu
20      25      30
Asp Arg Ser Leu Leu Glu Asp Ile Arg His Phe Asn Ile Phe Asp
35      40      45
Ser Val Leu Asp Leu Trp Ala Ala Cys Leu Tyr Arg Ser Cys Leu

```


Leu Leu Gly Ala Thr	50	Lys Asn Ser Ala Leu Gly	60
Pro Arg Arg Leu Arg Ala Ser Trp Leu Val Ile Ser Leu Val Cys	65		75
Leu Phe Val Gly Ile Tyr Ala Met Val Lys Leu Leu Leu Phe Ser	80		90
Glu Val Arg Arg Pro Ile Arg Asp Pro Trp Phe Trp Ala Leu Phe	95		105
Val Trp Thr Tyr Ile Ser Leu Gly Ala Ser Phe Leu Leu Trp Trp	110		120
Leu Leu Ser Thr Val Arg Pro Gly Thr Gln Ala Leu Glu Pro Gly	125		135
Ala Ala Thr Glu Ala Glu Gly Phe Pro Gly Ser Gly Arg Pro Pro	140		150
Pro Glu Gln Ala Ser Gly Ala Thr Leu Gln Lys Leu Leu Ser Tyr	155		165
Thr Lys Pro Asp Val Ala Phe Leu Val Ala Ala Ser Phe Phe Leu	170		180
Ile Val Ala Ala Leu Gly Glu Thr Phe Leu Pro Tyr Tyr Thr Gly	185		195
Arg Ala Ile Asp Gly Ile Val Ile Gln Lys Ser Met Asp Gln Phe	200		210
Ser Thr Ala Val Val Ile Val Cys Leu Leu Ala Ile Gly Ser Ser	215		225
Phe Ala Ala Gly Ile Arg Gly Gly Ile Phe Thr Leu Ile Phe Ala	230		240
Arg Leu Asn Ile Arg Leu Arg Asn Cys Leu Phe Arg Ser Leu Val	245		255
Ser Gln Glu Thr Ser Phe Phe Asp Glu Asn Arg Thr Gly Asp Leu	260		270
Ile Ser Arg Leu Thr Ser Asp Thr Thr Met Val Ser Asp Leu Val	275		285
Ser Gln Asn Ile Asn Val Phe Leu Arg Asn Thr Val Lys Val Thr	290		300
Gly Val Val Val Phe Met Phe Ser Leu Ser Trp Gln Leu Ser Leu	305		315
Val Thr Phe Met Gly Phe Pro Ile Ile Met Met Val Ser Asn Ile	320		330
Tyr Gly Lys Tyr Tyr Lys Arg Leu Ser Lys Glu Val Gln Asn Ala	335		345
Leu Ala Arg Ala Ser Asn Thr Ala Glu Glu Thr Ile Ser Ala Met	350		360
Lys Thr Val Arg Ser Phe Ala Asn Glu Glu Glu Glu Ala Glu Val	365		375
Tyr Leu Arg Lys Leu Gln Gln Val Tyr Lys Leu Asn Arg Lys Glu	380		390
Ala Ala Ala Tyr Met Tyr Tyr Val Trp Gly Ser Gly Ser Val Gly	395		405
Ser Val Tyr Ser Gly Leu Met Gln Gly Val Gly Ala Ala Glu Lys	410		420
Val Phe Glu Phe Ile Asp Arg Gln Pro Thr Met Val His Asp Gly	425		435
Ser Leu Ala Pro Asp His Leu Glu Gly Arg Val Asp Phe Glu Asn	440		450
Val Thr Phe Thr Tyr Arg Thr Arg Pro His Thr Gln Val Leu Gln	455		465
Asn Val Ser Phe Ser Leu Ser Pro Gly Lys Val Thr Ala Leu Val	470		480
Gly Pro Ser Gly Ser Gly Lys Ser Ser Cys Val Asn Ile Leu Glu	485		495
Asn Phe Tyr Pro Leu Glu Gly Gly Arg Val Leu Leu Asp Gly Lys	500		510
	515		525

```

Pro Ile Ser Ala Tyr Asp His Lys Tyr Leu His Arg Val Ile Ser
530 535
Leu Val Ser Gln Glu Pro Val Leu Phe Ala Arg Ser Ile Thr Asp
545 550
Asn Ile Ser Tyr Gly Leu Pro Thr Val Pro Phe Glu Met Val Val
560 565
Glu Ala Ala Gln Lys Ala Asn Ala His Gly Phe Ile Met Glu Leu
575 580
Gln Asp Gly Tyr Ser Thr Glu Thr Gly Glu Lys Gly Ala Gln Leu
590 595
Ser Gly Gly Gln Lys Gln Arg Val Ala Met Ala Arg Ala Leu Val
605 610
Arg Asn Pro Pro Val Leu Ile Leu Asp Glu Ala Thr Ser Ala Leu
620 625
Asp Ala Glu Ser Glu Tyr Leu Ile Gln Gln Ala Ile His Gly Asn
635 640
Leu Gln Lys His Thr Val Leu Ile Ile Ala His Arg Leu Ser Thr
650 655
Val Glu His Ala His Leu Ile Val Val Leu Asp Lys Gly Arg Val
665 670
Val Gln Gln Gly Thr His Gln Gln Leu Leu Ala Gln Gly Gly Leu
680 685
Tyr Ala Lys Leu Val Gln Arg Gln Met Leu Gly Leu Gln Pro Ala
695 700
Ala Asp Phe Thr Ala Gly His Asn Glu Pro Val Ala Asn Gly Ser
710 715
His Lys Ala

```

<210> 5

<211> 1009

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 5151730CD1

<400> 5

```

Met Glu Ala Leu Thr Leu Trp Leu Leu Pro Trp Ile Cys Gln Cys
1 5 10 15
Val Ser Val Arg Ala Asp Ser Ile Ile His Ile Gly Ala Ile Phe
20 25 30
Glu Glu Asn Ala Ala Lys Asp Asp Arg Val Phe Gln Leu Ala Val
35 40 45
Ser Asp Leu Ser Leu Asn Asp Asp Ile Leu Gln Ser Glu Lys Ile
50 55 60
Thr Tyr Ser Ile Lys Val Ile Glu Ala Asn Asn Pro Phe Gln Ala
65 70 75
Val Gln Glu Ala Cys Asp Leu Met Thr Gln Gly Ile Leu Ala Leu
80 85 90
Val Thr Ser Thr Gly Cys Ala Ser Ala Asn Ala Leu Gln Ser Leu
95 100 105
Thr Asp Ala Met His Ile Pro His Leu Phe Val Gln Arg Asn Pro
110 115 120
Gly Gly Ser Pro Arg Thr Ala Cys His Leu Asn Pro Ser Pro Asp
125 130 135
Gly Glu Ala Tyr Thr Leu Ala Ser Arg Pro Pro Val Arg Leu Asn
140 145 150
Asp Val Met Leu Arg Leu Val Thr Glu Leu Arg Trp Gln Lys Phe
155 160 165
Val Met Phe Tyr Asp Ser Glu Tyr Asp Ile Arg Gly Leu Gln Ser
170 175 180

```

Phe Leu Asp Gln Ala Ser Arg Leu Gly Leu Asp Val Ser Leu Gln
 185 190 195
 Lys Val Asp Lys Asn Ile Ser His Val Phe Thr Ser Leu Phe Thr
 200 205 210
 Thr Met Lys Thr Glu Glu Leu Asn Arg Tyr Arg Asp Thr Leu Arg
 215 220 225
 Arg Ala Ile Leu Leu Leu Ser Pro Gln Gly Ala His Ser Phe Ile
 230 235 240
 Asn Glu Ala Val Glu Thr Asn Leu Ala Ser Lys Asp Ser His Trp
 245 250 255
 Val Phe Val Asn Glu Glu Ile Ser Asp Pro Glu Ile Leu Asp Leu
 260 265 270
 Val His Ser Ala Leu Gly Arg Met Thr Val Val Arg Gln Ile Phe
 275 280 285
 Pro Ser Ala Lys Asp Asn Gln Lys Cys Thr Arg Asn Asn His Arg
 290 295 300
 Ile Ser Ser Leu Leu Cys Asp Pro Gln Glu Gly Tyr Leu Gln Met
 305 310 315
 Leu Gln Ile Ser Asn Leu Tyr Leu Tyr Asp Ser Val Leu Met Leu
 320 325 330
 Ala Asn Ala Phe His Arg Lys Leu Glu Asp Arg Lys Trp His Ser
 335 340 345
 Met Ala Ser Leu Asn Cys Ile Arg Lys Ser Thr Lys Pro Trp Asn
 350 355 360
 Gly Gly Arg Ser Met Leu Asp Thr Ile Lys Lys Gly His Ile Thr
 365 370 375
 Gly Leu Thr Gly Val Met Glu Phe Arg Glu Asp Ser Ser Asn Pro
 380 385 390
 Tyr Val Gln Phe Glu Ile Leu Gly Thr Thr Tyr Ser Glu Thr Phe
 395 400 405
 Gly Lys Asp Met Arg Lys Leu Ala Thr Trp Asp Ser Glu Lys Gly
 410 415 420
 Leu Asn Gly Ser Leu Gln Glu Arg Pro Met Gly Ser Arg Leu Gln
 425 430 435
 Gly Leu Thr Leu Lys Val Val Thr Val Leu Glu Glu Pro Phe Val
 440 445 450
 Met Val Ala Glu Asn Ile Leu Gly Gln Pro Lys Arg Tyr Lys Gly
 455 460 465
 Phe Ser Ile Asp Val Leu Asp Ala Leu Ala Lys Ala Leu Gly Phe
 470 475 480
 Lys Tyr Glu Ile Tyr Gln Ala Pro Asp Gly Arg Tyr Gly His Gln
 485 490 495
 Leu His Asn Thr Ser Trp Asn Gly Met Ile Gly Glu Leu Ile Ser
 500 505 510
 Lys Arg Ala Asp Leu Ala Ile Ser Ala Ile Thr Ile Thr Pro Glu
 515 520 525
 Arg Glu Ser Val Val Asp Phe Ser Lys Arg Tyr Met Asp Tyr Ser
 530 535 540
 Val Gly Ile Leu Ile Lys Lys Pro Glu Glu Lys Ile Ser Ile Phe
 545 550 555
 Ser Leu Phe Ala Pro Phe Asp Phe Ala Val Trp Ala Cys Ile Ala
 560 565 570
 Ala Ala Ile Pro Val Val Gly Val Leu Ile Phe Val Leu Asn Arg
 575 580 585
 Ile Gln Ala Val Arg Ala Gln Ser Ala Ala Gln Pro Arg Pro Ser
 590 595 600
 Ala Ser Ala Thr Leu His Ser Ala Ile Trp Ile Val Tyr Gly Ala
 605 610 615
 Phe Val Gln Gln Gly Gly Glu Ser Ser Val Asn Ser Met Ala Met
 620 625 630
 Arg Ile Val Met Gly Ser Trp Trp Leu Phe Thr Leu Ile Val Cys
 635 640 645
 Ser Ser Tyr Thr Ala Asn Leu Ala Ala Phe Leu Thr Val Ser Arg

Met Asp Asn Pro	650	Arg Thr Phe Gln	655	Leu Ser Lys Gln Val	660
	665		670		675
Glu Met Ser Tyr	680	Thr Val Arg Asp	685	Ser Ala Val Tyr Glu Tyr	690
	695		700		705
Phe Arg Ala Lys	710	Thr Asn Pro Leu	715	Glu Gln Asp Ser Thr Phe	720
	725		730		735
Ala Glu Leu Trp	740	Arg Thr Ile Ser Lys	745	Asn Gly Gly Ala Asp Asn	750
	755		760		765
Cys Val Ser Ser	770	Pro Ser Glu Gly Ile	775	Arg Lys Ala Lys Lys Gly	780
	785		790		795
Asn Tyr Ala Phe	800	Leu Trp Asp Val Ala	805	Val Val Glu Tyr Ala Ser	810
	815		820		825
Leu Thr Asp Asp	830	Asp Cys Ser Val Thr	835	Val Ile Gly Asn Ser Ile	840
	845		850		855
Ser Ser Lys Gly	860	Tyr Gly Ile Ala Leu	865	Gln His Gly Ser Pro Tyr	870
	875		880		885
Arg Asp Leu Phe	890	Ser Gln Arg Ile Leu	895	Glu Leu Gln Asp Thr Gly	900
	905		910		915
Asp Leu Asp Val	920	Leu Lys Gln Lys Trp	925	Trp Pro His Met Gly Arg	930
	935		940		945
Cys Asp Leu Thr	950	Ser His Ala Ser Ala	955	Gln Ala Asp Gly Lys Ser	960
	965		970		975
Leu Lys Leu His	980	Ser Phe Ala Gly Val	985	Phe Cys Ile Leu Ala Ile	990
	995		1000		1005
Gly Leu Leu Leu		Ala Cys Leu Val Ala		Ala Leu Glu Leu Trp Trp	
Asn Ser Asn Arg		Cys His Gln Glu Thr		Pro Lys Glu Asp Lys Glu	
Val Asn Leu Glu		Gln Val His Arg Arg		Met Asn Ser Leu Met Asp	
Glu Asp Ile Ala		His Lys Gln Ile Ser		Pro Ala Ser Ile Glu Leu	
Ser Ala Leu Glu		Met Gly Gly Leu Ala		Pro Thr Gln Thr Leu Glu	
Pro Thr Arg Glu		Tyr Gln Asn Thr Gln		Leu Ser Val Ser Thr Phe	
Leu Pro Glu Gln		Ser Ser His Gly Thr		Ser Arg Thr Leu Ser Ser	
Gly Pro Ser Ser		Asn Leu Pro Leu Pro		Leu Ser Ser Ser Ala Thr	
Met Pro Ser Met		Gln Cys Lys His Arg		Ser Pro Asn Gly Gly Leu	
Phe Arg Gln Ser		Pro Val Lys Thr Pro		Ile Pro Met Ser Phe Gln	
Pro Val Pro Gly		Gly Val Leu Pro Glu		Ala Leu Asp Thr Ser His	
Gly Thr Ser Ile					

<210> 6

<211> 374

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7472584CD1

<400> 6

Met Lys Arg Gln Asn Val Arg Thr Leu Ser Leu Ile Val Cys Thr

1

5

10

15

Phe Thr Tyr Leu Leu Val Gly Ala Ala Val Phe Asp Ala Leu Glu

Ser Asp His Glu	20	Met Arg Glu Glu Glu	25	Leu Lys Ala Glu	30
Ile Arg Ile Lys	35	Gly Lys Tyr Asn Ile	40	Ser Ser Glu Asp Tyr	45
Gln Leu Glu Leu	50	Val Ile Leu Gln Ser	55	Pro His Arg Ala	60
Val Gln Trp Lys	65	Phe Ala Gly Ser Phe	70	Tyr Phe Ala Ile Thr	75
Ile Thr Thr Ile	80	Gly Tyr Gly His Ala	85	Pro Gly Thr Asp Ala	90
Gly Lys Ala Phe	95	Cys Met Phe Tyr Ala	100	Val Leu Gly Ile Pro	105
Thr Leu Val Met	110	Phe Gln Ser Leu Gly	115	Glu Arg Met Asn Thr	120
Val Arg Tyr Leu	125	Leu Lys Arg Ile Lys	130	Cys Cys Gly Met Arg	135
Asn Thr Asp Val	140	Ser Met Glu Asn Met	145	Val Thr Val Gly Phe	150
Ser Cys Met Gly	155	Thr Leu Cys Ile Gly	160	Ala Ala Phe Ser Gln	165
Cys Glu Glu Trp	170	Ser Phe Phe His Ala	175	Tyr Tyr Cys Phe Ile	180
Thr Leu Thr Thr	185	Ile Gly Phe Gly Asp	190	Tyr Val Ala Leu Gln	195
Lys Gly Ala Leu	200	Gln Lys Lys Pro Leu	205	Tyr Val Ala Phe Ser	210
Met Tyr Ile Leu	215	Val Gly Leu Thr Val	220	Ile Gly Ala Phe Leu	225
Leu Val Val Leu	230	Arg Phe Leu Thr Met	235	Asn Ser Glu Asp Glu	240
Arg Asp Ala Glu	245	Glu Arg Ala Ser Leu	250	Ala Gly Asn Arg Asn	255
Met Val Ile His	260	Ile Pro Glu Glu Pro	265	Arg Pro Ser Arg Pro	270
Tyr Lys Ala Asp	275	Val Pro Asp Leu Gln	280	Ser Val Cys Ser Cys	285
Cys Tyr Arg Ser	290	Gln Asp Tyr Gly Gly	295	Arg Ser Val Ala Pro	300
Asn Ser Phe Ser	305	Ala Lys Leu Ala Pro	310	His Tyr Phe His Ser	315
Ser Tyr Lys Ile	320	Glu Glu Ile Ser Pro	325	Thr Leu Lys Asn Ser	330
Leu Phe Pro Ser	335	Pro Ile Ser Ser Ile	340	Ser Pro Gly Leu His	345
Phe Thr Asp His	350	Gln Arg Leu Met Lys	355	Arg Arg Lys Ser Val	360
	365		370		

<210> 7

<211> 589

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7473536CD1

<400> 7

Met Ser Ala Val Leu	1	Thr Pro Gly Leu Phe	10	Leu Pro Leu Pro Gly	15
Pro Leu Pro Ala Ser	20	Leu His Lys Ala Gly	25	Gly Thr Gly Pro Gln	30
Val Arg Pro Met Ala	30	Met Ala Phe Thr Asp	35	Leu Leu Asp Ala Leu	40

	35		40		45
Gly Ser Met Gly	Arg Phe Gln Leu Asn His	Thr Ala Leu Leu Leu			
	50		55		60
Leu Pro Cys Gly	Leu Leu Ala Cys His Asn	Phe Leu Gln Asn Phe			
	65		70		75
Thr Ala Ala Val	Pro Pro His His Cys Arg	Gly Pro Ala Asn His			
	80		85		90
Thr Glu Ala Ser	Thr Asn Asp Ser Gly Ala	Trp Leu Arg Ala Thr			
	95		100		105
Ile Pro Leu Asp	Gln Leu Gly Ala Pro Glu	Pro Cys Arg Arg Phe			
	110		115		120
Thr Lys Pro Gln	Trp Ala Leu Leu Ser Pro	Asn Ser Ser Ile Pro			
	125		130		135
Gly Ala Ala Thr	Glu Gly Cys Lys Asp Gly	Trp Val Tyr Asn Arg			
	140		145		150
Ser Val Phe Pro	Ser Thr Ile Val Met Glu	Gln Trp Asp Leu Val			
	155		160		165
Cys Glu Ala Arg	Thr Leu Arg Asp Leu Ala	Gln Ser Val Tyr Ile			
	170		175		180
Ala Gly Val Leu	Val Gly Ala Ala Val Phe	Gly Ser Leu Ala Asp			
	185		190		195
Arg Leu Gly Cys	Lys Gly Pro Leu Val Trp	Ser Tyr Leu Gln Leu			
	200		205		210
Ala Ala Ser Gly	Ala Ala Thr Ala Tyr Phe	Ser Ser Phe Ser Ala			
	215		220		225
Tyr Cys Val Phe	Arg Phe Leu Met Gly Met	Thr Phe Ser Gly Ile			
	230		235		240
Ile Leu Asn Ser	Val Ser Leu Val Glu Trp	Met Pro Thr Arg Gly			
	245		250		255
Arg Thr Val Ala	Gly Ile Leu Leu Gly Tyr	Ser Phe Thr Leu Gly			
	260		265		270
Gln Leu Ile Leu	Ala Gly Val Ala Tyr Leu	Ile Arg Pro Trp Arg			
	275		280		285
Cys Leu Gln Phe	Ala Ile Ser Ala Pro Phe	Leu Ile Phe Phe Leu			
	290		295		300
Tyr Ser Trp Trp	Leu Pro Glu Ser Ser Arg	Trp Leu Leu Leu His			
	305		310		315
Gly Lys Ser Gln	Leu Ala Val Gln Asn Leu	Gln Lys Val Ala Ala			
	320		325		330
Met Asn Gly Arg	Lys Gln Glu Gly Glu Arg	Leu Thr Lys Glu Val			
	335		340		345
Met Ser Ser Tyr	Ile Gln Ser Glu Phe Ala	Ser Val Cys Thr Ser			
	350		355		360
Asn Ser Ile Leu	Asp Leu Phe Arg Thr Pro	Ala Ile Arg Lys Val			
	365		370		375
Thr Cys Cys Pro	Ala Leu Arg Phe Ser Asn	Ser Val Ala Tyr Tyr			
	380		385		390
Gly Leu Ala Met	Asp Leu Gln Lys Phe Gly	Leu Ser Leu Tyr Leu			
	395		400		405
Val Gln Ala Leu	Phe Gly Ile Ile Asn Ile	Pro Ala Met Leu Val			
	410		415		420
Ala Thr Ala Thr	Met Ile Tyr Val Gly Arg	Arg Ala Thr Val Ala			
	425		430		435
Ser Phe Leu Ile	Leu Ala Gly Leu Met Val	Ile Ala Asn Met Phe			
	440		445		450
Val Pro Glu Gly	Thr Gln Ile Leu Cys Thr	Ala Gln Ala Ala Leu			
	455		460		465
Gly Lys Gly Cys	Leu Ala Ser Ser Phe Ile	Cys Val Tyr Leu Phe			
	470		475		480
Thr Gly Glu Leu	Tyr Pro Thr Glu Ile Arg	Gln Met Gly Met Gly			
	485		490		495
Phe Ala Ser Val	His Ala Arg Leu Gly Gly	Leu Thr Ala Pro Leu			
	500		505		510

```

Val Thr Thr Leu Gly Glu Tyr Ser Thr Ile Leu Pro Pro Val Ser
      515      520
Phe Gly Ala Thr Ala Ile Leu Ala Gly Leu Ala Val Cys Phe Leu
      530      535
Thr Glu Thr Arg Asn Met Pro Leu Val Glu Thr Ile Ala Ala Met
      545      550
Glu Arg Arg Val Lys Glu Gly Ser Ser Lys Lys His Val Glu Glu
      560      565
Lys Ser Glu Glu Ile Ser Leu Gln Gln Leu Arg Ala Ser Pro Leu
      575      580
Lys Glu Thr Ile
      585

```

```

<210> 8
<211> 549
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7473422CD1

```

```

<400> 8
Met Val Ser Asp Arg Gly Leu Lys Pro Phe Glu Asp Leu Arg Pro
  1      5      10
Pro Lys Ile Ser Pro Ala Asp Leu Gly Asn Ala Glu Glu Ala Ile
  20      25      30
Glu Leu Glu Arg Glu Pro Ala Pro Val Arg Phe Val Pro Arg Arg
  35      40      45
Lys Arg Pro Ser Ile Trp Val Val Leu Pro Val Leu Phe Leu Val
  50      55      60
Ala Met Ser Leu Leu Pro Leu Leu Tyr Val Ala Ile Lys Ala Trp
  65      70      75
Glu Ala Gly Trp Arg Glu Ala Leu His Leu Leu Trp Arg Pro Phe
  80      85      90
Val Trp Gly Leu Met Arg Asn Thr Leu Met Leu Met Val Gly Val
  95      100      105
Thr Leu Ala Cys Met Val Val Gly Leu Ala Leu Ala Trp Leu Leu
  110      115      120
Glu Arg Ser Asn Leu Ala Gly Arg Arg Leu Trp Gly Val Val Leu
  125      130      135
Cys Leu Pro Phe Ala Val Pro Ser Phe Val Ser Ser Phe Thr Trp
  140      145      150
Val Ser Leu Ser Ser Asp Phe Glu Gly Leu Gly Gly Ala Ile Leu
  155      160      165
Val Met Ala Leu Ser Lys Tyr Pro Leu Val Phe Leu Pro Val Ala
  170      175      180
Ala Thr Leu Arg Asn Leu Asp Thr Ser Leu Glu Glu Ser Ala Arg
  185      190      195
Thr Leu Gly Cys Ser Arg Trp Gly Val Phe Ile Lys Val Thr Leu
  200      205      210
Pro Leu Leu Trp Pro Ser Met Leu Gly Gly Ala Leu Leu Ile Ala
  215      220      225
Leu His Met Leu Val Glu Phe Gly Ala Leu Ser Ile Leu Gly Leu
  230      235      240
Gln Thr Phe Thr Thr Ala Ile Tyr Gln Gln Phe Glu Leu Glu Phe
  245      250      255
Ser Asn Ala Asn Ala Ala Met Leu Ser Ala Val Leu Leu Ala Met
  260      265      270
Cys Leu Val Met Leu Trp Leu Glu Leu Arg Val Arg Gly Lys Ala
  275      280      285
Arg His Val Arg Ile Gly Gln Gly Val Ala Arg Arg Ala Gln Pro
  290      295      300

```

Val Arg Leu Arg Gly Trp Ala Val Pro Ala Gln Leu Leu Cys Val
 305 310 315
 Ala Leu Ala Val Leu Gly Ser Gly Ile Pro Leu Ala Met Leu Gly
 320 325 330
 Tyr Trp Leu Ser Val Gly Ser Ser Ala Ala Phe Pro Val Gly Ala
 335 340 345
 Ile Ser Lys Ala Leu Phe Thr Ser Leu Ser Val Ser Leu Gly Gly
 350 355 360
 Ala Gly Phe Cys Val Leu Leu Ala Leu Pro Ile Ser Phe Leu Val
 365 370 375
 Val Arg Tyr Lys Gly Arg Leu Ala Ile Trp Ala Glu Arg Leu Pro
 380 385 390
 Tyr Leu Leu His Ala Leu Pro Gly Leu Val Ile Ala Leu Thr Leu
 395 400 405
 Val Phe Phe Ala Leu His Tyr Val Pro Ala Leu Tyr Gln Thr Thr
 410 415 420
 Ala Leu Leu Leu Leu Ala Tyr Ala Leu Leu Phe Leu Pro Leu Ala
 425 430 435
 Gln Ser Pro Val Arg Thr Ala Leu Asn Lys Ala Ser Pro Thr Leu
 440 445 450
 Glu Glu Ala Ala Arg Thr Leu Gly Ala Ser Ser Phe Thr Ala Phe
 455 460 465
 Cys Arg Val Thr Leu Pro Ile Ile Phe Pro Ala Met Ala Ala Ala
 470 475 480
 Phe Ala Leu Val Phe Leu Asp Ala Met Lys Glu Leu Thr Ala Thr
 485 490 495
 Leu Leu Leu Ser Pro Thr Gly Met Thr Thr Leu Ala Thr Glu Val
 500 505 510
 Trp Ala His Thr Ala Asn Val Glu Phe Ala Ala Ala Ala Pro Tyr
 515 520 525
 Ala Ala Leu Leu Ile Val Val Ser Gly Leu Pro Val Tyr Leu Leu
 530 535 540
 Thr Thr Arg Met Tyr Leu Asn Lys Ala
 545

<210> 9

<211> 634

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 2864715CD1

<400> 9

Met Val Arg Leu Val Leu Pro Asn Pro Gly Leu Asp Ala Arg Ile
 1 5 10 15
 Pro Ser Leu Ala Glu Leu Glu Thr Ile Glu Gln Glu Glu Ala Ser
 20 25 30
 Ser Arg Pro Lys Trp Asp Asn Lys Ala Gln Tyr Met Leu Thr Cys
 35 40 45
 Leu Gly Phe Cys Val Gly Leu Gly Asn Val Trp Arg Phe Pro Tyr
 50 55 60
 Leu Cys Gln Ser His Gly Gly Gly Ala Phe Met Ile Pro Phe Leu
 65 70 75
 Ile Leu Leu Val Leu Glu Gly Ile Pro Leu Leu Tyr Leu Glu Phe
 80 85 90
 Ala Ile Gly Gln Arg Leu Arg Arg Gly Ser Leu Gly Val Trp Ser
 95 100 105
 Ser Ile His Pro Ala Leu Lys Gly Leu Gly Leu Ala Ser Met Leu
 110 115 120
 Thr Ser Phe Met Val Gly Leu Tyr Tyr Asn Thr Ile Ile Ser Trp
 125 130 135

Ile Met Trp Tyr Leu Phe Asn Ser Phe Gln Glu Pro Leu Pro Trp
 140 150
 Ser Asp Trp Pro Leu Asn Glu Asn Gln Thr Gly Tyr Val Asp Glu
 155 165
 Cys Ala Arg Ser Ser Pro Val Asp Tyr Phe Trp Tyr Arg Glu Thr
 170 180
 Leu Asn Ile Ser Thr Ser Ile Ser Asp Ser Gly Ser Ile Gln Trp
 185 195
 Trp Met Leu Leu Cys Leu Ala Cys Ala Trp Ser Val Leu Tyr Met
 200 210
 Cys Thr Ile Arg Gly Ile Glu Thr Thr Gly Lys Ala Val Tyr Ile
 215 225
 Thr Ser Thr Leu Pro Tyr Val Val Leu Thr Ile Phe Leu Ile Arg
 230 240
 Gly Leu Thr Leu Lys Gly Ala Thr Asn Gly Ile Val Phe Leu Phe
 245 255
 Thr Pro Asn Val Thr Glu Leu Ala Gln Pro Asp Thr Trp Leu Asp
 260 270
 Ala Gly Ala Gln Val Phe Phe Ser Phe Ser Leu Ala Phe Gly Gly
 275 285
 Leu Ile Ser Phe Ser Ser Tyr Asn Ser Val His Asn Asn Cys Glu
 290 300
 Lys Asp Ser Val Ile Val Ser Ile Ile Asn Gly Phe Thr Ser Val
 305 315
 Tyr Val Ala Ile Val Val Tyr Ser Val Ile Gly Phe Arg Ala Thr
 320 330
 Gln Arg Tyr Asp Asp Cys Phe Ser Thr Asn Ile Leu Thr Leu Ile
 335 345
 Asn Gly Phe Asp Leu Pro Glu Gly Asn Val Thr Gln Glu Asn Phe
 350 360
 Val Asp Met Gln Gln Arg Cys Asn Ala Ser Asp Pro Ala Ala Tyr
 365 375
 Ala Gln Leu Val Phe Gln Thr Cys Asp Ile Asn Ala Phe Leu Ser
 380 390
 Glu Ala Val Glu Gly Thr Gly Leu Ala Phe Ile Val Phe Thr Glu
 395 405
 Ala Ile Thr Lys Met Pro Leu Ser Pro Leu Trp Ser Val Leu Phe
 410 420
 Phe Ile Met Leu Phe Cys Leu Gly Leu Ser Ser Met Phe Gly Asn
 425 435
 Met Glu Gly Val Val Val Pro Leu Gln Asp Leu Arg Val Ile Pro
 440 450
 Pro Lys Trp Pro Lys Glu Val Leu Thr Gly Leu Ile Cys Leu Gly
 455 465
 Thr Phe Leu Ile Gly Phe Ile Phe Thr Ala Gly Ser Ile Pro Leu Leu
 470 480
 Trp Phe Ser Leu Leu Asp Ser Tyr Ala Gly Ser Ile Pro Leu Leu
 485 495
 Ile Ile Ala Phe Cys Glu Met Phe Ser Val Val Tyr Val Tyr Gly
 500 510
 Val Asp Arg Phe Asn Lys Asp Ile Glu Phe Met Ile Gly His Lys
 515 525
 Pro Asn Ile Phe Trp Gln Val Thr Trp Arg Val Val Ser Pro Leu
 530 540
 Leu Met Leu Ile Ile Phe Leu Phe Phe Phe Val Val Glu Val Ser
 545 555
 Gln Glu Leu Thr Tyr Ser Ile Trp Asp Pro Gly Tyr Glu Glu Phe
 560 570
 Pro Lys Ser Gln Lys Ile Ser Tyr Pro Asn Trp Val Tyr Val Val
 575 585
 Val Val Ile Val Ala Gly Val Pro Ser Leu Thr Ile Pro Gly Tyr
 590 600
 Ala Ile Tyr Lys Leu Ile Arg Asn His Cys Gln Lys Pro Gly Asp

	605		610		615
His Gln Gly Leu	Val Ser Thr Leu Ser	Thr Ala Ser Met Asn Gly			
	620	625		630	
Asp Leu Lys Tyr					

<210> 10
 <211> 491
 <212> PRT
 <213> Homo sapiens

<220>
 <221> misc_feature
 <223> Incyte ID No: 1734724CD1

<400> 10
 Met Asp Gly Asn Asp Asn Val Thr Leu Leu Phe Ala Pro Leu Leu
 1 5 10 15
 Arg Asp Asn Tyr Thr Leu Ala Pro Asn Ala Ser Ser Leu Gly Pro
 20 25 30
 Gly Thr Asn Leu Ala Leu Ala Pro Ala Ser Ser Ala Gly Pro Gly
 35 40 45
 Pro Gly Leu Ser Leu Gly Pro Val Pro Ser Phe Gly Phe Ser Pro
 50 55 60
 Gly Pro Thr Pro Thr Pro Glu Pro Thr Thr Ser Gly Leu Ala Gly
 65 70 75
 Gly Ala Ala Ser His Gly Pro Ser Pro Val Pro Ser Ala Leu Gly
 80 85 90
 Ala Pro Arg Ala Pro Val Leu Gly His Ala Ala Glu Pro Arg Ala
 95 100 105
 Glu Arg Val Arg Gly Arg Arg Leu Cys Ile Thr Met Leu Gly Leu
 110 115 120
 Gly Cys Thr Val Asp Val Asn His Phe Gly Ala His Val Arg Arg
 125 130 135
 Pro Val Ala Ala Leu Leu Ala Ala Leu Pro Val Arg Pro Pro Ala
 140 145 150
 Ala Ala Gly Leu Pro Ala Gly Pro Arg Leu Gln Ala Gly Arg Gly
 155 160 165
 Gly Arg Arg Gly Leu Leu Cys Gly Cys Cys Pro Gly Gly Asn
 170 175 180
 Leu Ser Asn Leu Met Ser Leu Leu Val Asp Gly Asp Met Asn Leu
 185 190 195
 Arg Arg Ala Ala Leu Leu Ala Leu Ser Ser Asp Val Gly Ser Ala
 200 205 210
 Gln Thr Ser Thr Pro Gly Leu Ala Val Ser Pro Phe His Leu Tyr
 215 220 225
 Ser Thr Tyr Lys Lys Val Ser Trp Leu Phe Asp Ser Lys Leu
 230 235 240
 Val Leu Ile Ser Ala His Ser Leu Phe Cys Ser Ile Ile Met Thr
 245 250 255
 Ile Ser Ser Thr Leu Leu Ala Leu Val Leu Met Pro Leu Cys Leu
 260 265 270
 Trp Ile Tyr Ser Trp Ala Trp Ile Asn Thr Pro Ile Val Gln Leu
 275 280 285
 Leu Pro Leu Gly Thr Val Thr Leu Thr Leu Cys Ser Thr Leu Ile
 290 295 300
 Pro Ile Gly Leu Gly Val Phe Ile Arg Tyr Lys Tyr Ser Arg Val
 305 310 315
 Ala Asp Tyr Ile Val Lys Val Ser Leu Trp Ser Leu Leu Val Thr
 320 325 330
 Leu Val Val Leu Phe Ile Met Thr Gly Thr Met Leu Gly Pro Glu
 335 340 345
 Leu Leu Ala Ser Ile Pro Ala Ala Val Tyr Val Ile Ala Ile Phe

	350		355		360
Met	Pro	Leu	Ala	Gly	Tyr
	365		370		375
Phe	His	Leu	Pro	Pro	Asn
	380		385		390
Gly	Ser	Gln	Asn	Val	Gln
	395		400		405
Phe	Pro	Pro	Gln	Phe	Ile
	410		415		420
Tyr	Ala	Leu	Phe	Gln	Ser
	425		430		435
Tyr	Lys	Met	Tyr	Gly	Ser
	440		445		450
Asp	Glu	Asp	Glu	Asp	Thr
	455		460		465
Glu	Glu	Met	Ala	Asp	Thr
	470		475		480
Ile	Ile	Met	Met	Glu	Thr
	485		490		

<210> 11

<211> 525

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 1563237CD1

<400> 11

Met	Pro	Ala	Pro	Arg	Ala	Arg	Glu	Gln	Pro	Arg	Val	Pro	Gly	Glu
1	5								10					15
Arg	Gln	Pro	Leu	Leu	Pro	Arg	Gly	Ala	Arg	Gly	Pro	Arg	Arg	Trp
			20						25					30
Arg	Arg	Ala	Ala	Gly	Ala	Ala	Val	Leu	Leu	Val	Glu	Met	Leu	Glu
			35						40					45
Arg	Ala	Ala	Phe	Phe	Gly	Val	Thr	Ala	Asn	Leu	Val	Leu	Tyr	Leu
			50						55					60
Asn	Ser	Thr	Asn	Phe	Asn	Trp	Thr	Gly	Glu	Gln	Ala	Thr	Arg	Ala
			65						70					75
Ala	Leu	Val	Phe	Leu	Gly	Ala	Ser	Tyr	Leu	Leu	Ala	Pro	Val	Gly
			80						85					90
Gly	Trp	Leu	Ala	Asp	Val	Tyr	Leu	Gly	Arg	Tyr	Arg	Ala	Val	Ala
			95						100					105
Leu	Ser	Leu	Leu	Leu	Tyr	Leu	Ala	Ala	Ser	Gly	Leu	Leu	Pro	Ala
			110						115					120
Thr	Ala	Phe	Pro	Asp	Gly	Arg	Ser	Ser	Phe	Cys	Gly	Glu	Met	Pro
			125						130					135
Ala	Ser	Pro	Leu	Gly	Pro	Ala	Cys	Pro	Ser	Ala	Gly	Cys	Pro	Arg
			140						145					150
Ser	Ser	Pro	Ser	Pro	Tyr	Cys	Ala	Pro	Val	Leu	Tyr	Ala	Gly	Leu
			155						160					165
Leu	Leu	Leu	Gly	Leu	Ala	Ala	Ser	Ser	Val	Arg	Ser	Asn	Leu	Thr
			170						175					180
Ser	Phe	Gly	Ala	Asp	Gln	Val	Met	Asp	Leu	Gly	Arg	Asp	Ala	Thr
			185						190					195
Arg	Arg	Phe	Phe	Asn	Trp	Phe	Tyr	Trp	Ser	Ile	Asn	Leu	Gly	Ala
			200						205					210
Val	Leu	Ser	Leu	Leu	Val	Val	Ala	Phe	Ile	Gln	Gln	Asn	Ile	Ser
			215						220					225
Phe	Leu	Leu	Gly	Tyr	Ser	Ile	Pro	Val	Gly	Cys	Val	Gly	Leu	Ala
			230						235					240
Phe	Phe	Ile	Phe	Leu	Phe	Ala	Thr	Pro	Val	Phe	Ile	Thr	Lys	Pro

Pro Met Gly Ser	245	250	255
Gln Val Ser Ser Met	260	265	270
Asn Cys Cys Pro	275	280	285
Gln Cys Ala Arg	290	295	300
Ala Ser Pro Gln	305	310	315
Ile Leu Pro Val	320	325	330
Phe Gln Met Gln	335	340	345
Ile Pro Asn Ile	350	355	360
Leu Arg Ala Gln	365	370	375
Ala Leu His Tyr	380	385	390
Gly Glu Val Leu	395	400	405
Ile Pro Gln Tyr	410	415	420
Ile Pro Gly Leu	425	430	435
Gln Gly Ala Ile	440	445	450
Ser Leu Leu Gly	455	460	465
Gly Trp Leu His	470	475	480
Arg Met Asp Leu	485	490	495
Thr Ala Leu Leu	500	505	510
Ser Gln Gly Pro	515	520	525

<210> 12

<211> 1310

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7473443CD1

<400> 12

Met Gly Lys Lys	Gln Cys Lys Lys	Ala Lys	Asn Ser Lys	Asn Gln
1	5	10	15	
Asn Ala Ser Ser	Pro Pro Lys Asp	His Asn	Ser Ser Pro	Ala Gly
20	25	30		
Glu Gln Asn Trp	Met Glu Asn Glu	Leu Thr	Glu Ala Gly	Phe Arg
35	40	45		
Arg Trp Val Val	Ile Asn Ser Cys	Lys Leu	Lys Glu His	Val Leu
50	55	60		
Thr Gln Cys Lys	Glu Ala Lys Asn	Leu Glu	Lys Arg Leu	Gly Glu
65	70	75		
Leu Leu Thr Arg	Ile Thr Ser Leu	Glu Lys	Asn Ile Asn	Asp Leu
80	85	90		
Met Glu Leu Lys	Asn Thr Ala Arg	Glu Leu	Arg Asp Ala	Tyr Ile
95	100	105		

Ser Ile Ser Ser Arg Ile Asp Gln Ala Glu Lys Arg Ile Ser Glu
 110 115 120
 Ile Glu Asp Gln Leu Asn Glu Ile Lys Arg Glu Asp Lys Ile Arg
 125 130 135
 Glu Lys Asn Glu Lys Asp Glu Gln Gly Leu Gln Glu Ile Trp Asp
 140 145 150
 Tyr Val Lys Arg Pro Asn Leu His Leu Ile Gly Val Pro Gly Leu
 155 160 165
 Leu Tyr Ser Asp Met Cys Arg Leu Leu Pro Ser Pro Arg Asn Gln
 170 175 180
 Pro Ala Leu Gln Ala Leu Glu Arg Gly Val Ile Leu Glu Val Lys
 185 190 195
 Cys Val Val Cys Ser Thr Gln Ala Gly Ala Ala Arg Arg Gly Val
 200 205 210
 Lys Ile Ser Ile Lys Gly Lys Gly Phe Ser Val Val Ser Val Val
 215 220 225
 Gly Thr Leu Gln Trp Leu Leu Trp Ala Arg Ala Ala His Ala Pro
 230 235 240
 His Trp Arg Phe Leu Arg Trp Met Ala Ala Leu Trp Asp Val Pro
 245 250 255
 Gly Lys Thr Gly Pro Ser Pro Ile Ser Leu Thr Gly Gln Arg Gly
 260 265 270
 Asn Arg Gly Pro Glu Ser Ser Ser Ile Leu Arg Gly Val Pro Lys
 275 280 285
 Asp Phe Ser Thr Gly Thr Ser Ala Gln Leu Arg Arg Ala Met Gly
 290 295 300
 Leu Ala Pro Glu Gly Gly Gly Phe Gln Ala Phe Phe Pro Arg Pro
 305 310 315
 Thr Met Pro Ala Thr Pro Asn Phe Leu Ala Asn Pro Ser Ser Ser
 320 325 330
 Ser Arg Trp Ile Pro Leu Gln Pro Met Pro Val Ala Trp Ala Phe
 335 340 345
 Val Gln Lys Thr Ser Ala Leu Leu Trp Leu Leu Leu Gly Thr
 350 355 360
 Ser Leu Ser Pro Ala Trp Gly Gln Ala Lys Ile Pro Leu Glu Thr
 365 370 375
 Val Lys Leu Trp Ala Asp Thr Phe Gly Gly Asp Leu Tyr Asn Thr
 380 385 390
 Val Thr Lys Tyr Ser Gly Ser Leu Leu Leu Gln Lys Lys Tyr Lys
 395 400 405
 Asp Val Glu Ser Ser Leu Lys Ile Glu Glu Val Asp Gly Leu Glu
 410 415 420
 Leu Val Arg Lys Phe Ser Glu Asp Met Glu Asn Met Leu Arg Arg
 425 430 435
 Lys Val Glu Ala Val Gln Asn Leu Val Glu Ala Ala Glu Glu Ala
 440 445 450
 Asp Leu Asn His Glu Phe Asn Glu Ser Leu Val Glu Pro Gly Val
 455 460 465
 Gly Val Gly Val Gly Met Ser Val Thr Gln Ser Gly Val Gly Val
 470 475 480
 Gly Val Gly Met Ser Val Thr Gln Ser Gly Val Gly Val Gly Val
 485 490 495
 Gly Met Ser Ile Thr Leu Ser Gly Val Gly Val Gly Val Gly Met
 500 505 510
 Ser Val Arg Gln Ser Gly Val Gly Val Gly Val Gly Met Ser Val
 515 520 525
 Thr Gln Ser Gly Val Gly Val Gly Val Gly Met Ser Val Thr Gln
 530 535 540
 Ser Gly Val Gly Val Gly Val Gly Met Ser Val Arg Gln Ser Gly
 545 550 555
 Val Gly Val Gly Val Gly Met Ser Val Thr Gln Ser Trp Gly Val
 560 565 570
 Phe Ser Ala Gln Arg Ala Ala Ala Gly Ala Cys Val Asp Ser Asp

			575			580			585
Gly	Arg	Pro	Ala	Pro	Ala	Leu	Ser	Ser	His
			590						595
Ser	Ser	Ser	Leu	Ser	Ala	Cys	Pro	Gly	Ala
			605						610
Gly	Leu	Thr	Arg	Pro	Gln	Phe	Asp	Tyr	Tyr
			620						625
Ile	Asn	Glu	Arg	Asp	Glu	Lys	Gly	Asn	Phe
			635						640
Glu	Phe	Leu	Leu	Glu	Ser	Asn	Ala	His	Phe
			650						655
Asn	Thr	Ser	Ile	Ser	Ser	Val	Gln	Leu	Pro
			665						670
Lys	Asp	Pro	Asp	Ile	Leu	Asn	Gly	Val	Tyr
			680						685
Asn	Ala	Val	Phe	Val	Glu	Asn	Phe	Gln	Arg
			695						700
Trp	Gln	Tyr	Phe	Gly	Ser	Ala	Thr	Gly	Phe
			710						715
Gly	Ile	Lys	Trp	Thr	Pro	Asp	Glu	Asn	Gly
			725						730
Cys	Arg	Asn	Arg	Gly	Trp	Tyr	Ile	Gln	Ala
			740						745
Asp	Ile	Val	Ile	Leu	Val	Asp	Val	Ser	Gly
			755						760
Arg	Met	Thr	Ile	Ala	Lys	His	Thr	Ile	Thr
			770						775
Leu	Gly	Glu	Asn	Asp	Phe	Ile	Asn	Ile	Ile
			785						790
Val	His	Tyr	Ile	Glu	Pro	Cys	Phe	Lys	Gly
			800						805
Asp	Arg	Asp	Asn	Arg	Glu	His	Phe	Lys	Leu
			815						820
Met	Val	Lys	Gly	Val	Gly	Val	Val	Asp	Gln
			830						835
Phe	Gln	Ile	Leu	Lys	Gln	Phe	Gln	Glu	Ala
			845						850
Cys	Asn	Gln	Ala	Ile	Met	Leu	Ile	Ser	Asp
			860						865
Tyr	Glu	Pro	Val	Phe	Glu	Lys	Tyr	Asn	Trp
			875						880
Arg	Val	Phe	Thr	Tyr	Leu	Ile	Gly	Arg	Glu
			890						895
Arg	Met	Lys	Trp	Ile	Ala	Cys	Asn	Asn	Lys
			905						910
Ile	Ser	Thr	Leu	Ala	Asp	Thr	Gln	Glu	Asn
			920						925
His	Val	Leu	Ser	Arg	Pro	Met	Val	Ile	Asn
			935						940
Ile	Trp	Thr	Glu	Ala	Tyr	Met	Asp	Ser	Lys
			950						955
Ala	Gln	Ser	Leu	Thr	Leu	Leu	Thr	Thr	Val
			965						970
Ser	Lys	Lys	Asn	Glu	Thr	Arg	Ser	His	Gly
			980						985
Val	Gly	Ser	Asp	Val	Ala	Leu	Arg	Glu	Leu
			995						1000
Arg	Tyr	Lys	Leu	Gly	Val	His	Gly	Tyr	Ala
			1010						1015
Asn	Gly	Tyr	Ile	Leu	Ser	His	Pro	Asp	Leu
			1025						1030
Glu	Gly	Lys	Lys	Leu	Lys	Pro	Lys	Pro	Asn
			1040						1045
									1050

Leu Ser Glu Val Glu Trp Glu Asp Gln Ala Glu Ser Leu Arg Thr
 1055 1060 1065
 Ala Met Ile Asn Arg Glu Thr Gly Thr Leu Ser Met Asp Val Lys
 1070 1075 1080
 Val Pro Met Asp Lys Gly Lys Arg Val Leu Phe Leu Thr Asn Asp
 1085 1090 1095
 Tyr Phe Phe Thr Asp Ile Ser Asp Thr Pro Phe Ser Leu Gly Val
 1100 1105 1110
 Val Leu Ser Arg Gly His Gly Glu Tyr Ile Leu Leu Gly Asn Thr
 1115 1120 1125
 Ser Val Glu Glu Gly Leu His Asp Leu Leu His Pro Asp Leu Ala
 1130 1135 1140
 Leu Ala Gly Asp Trp Ile Tyr Cys Ile Thr Asp Ile Asp Pro Asp
 1145 1150 1155
 His Arg Lys Leu Ser Gln Leu Glu Ala Met Ile Arg Phe Leu Thr
 1160 1165 1170
 Arg Lys Asp Pro Asp Leu Glu Cys Asp Gly Glu Leu Val Arg Glu
 1175 1180 1185
 Val Leu Phe Asp Ala Val Val Thr Ala Pro Met Glu Ala Tyr Trp
 1190 1195 1200
 Thr Ala Leu Ala Leu Asn Met Ser Glu Glu Ser Glu His Val Val
 1205 1210 1215
 Asp Met Ala Phe Leu Gly Thr Arg Ala Gly Leu Leu Arg Ser Ser
 1220 1225 1230
 Leu Phe Val Gly Ser Glu Lys Val Ser Asp Arg Lys Phe Leu Thr
 1235 1240 1245
 Pro Glu Asp Glu Ala Ser Val Phe Thr Leu Asp Arg Phe Pro Leu
 1250 1255 1260
 Trp Tyr Arg Gln Ala Ser Glu His Pro Ala Gly Ser Phe Val Phe
 1265 1270 1275
 Asn Leu Arg Trp Ala Glu Gly Pro Gly Arg Pro Ser Ala Lys Gly
 1280 1285 1290
 Leu Pro Pro Pro Leu Cys Gln Thr Ile Leu Lys Arg Arg Asp Gly
 1295 1300 1305
 Lys Met Ser Trp Ser
 1310

<210> 13
 <211> 400
 <212> PRT
 <213> Homo sapiens

 <220>
 <221> misc_feature
 <223> Incyte ID No: 7473438CD1

<400> 13
 Met Asn Pro Gly Gln Ala Ser Gly Arg Arg Thr Gly Glu Arg Phe
 1 5 10 15
 Phe Arg Pro Pro Pro Val Ala Ile Pro Ala Ser Arg Phe Pro Ala
 20 25 30
 Val Ala Pro Pro Arg Pro Ser Gln Pro Cys Arg Val Gly Pro Gly
 35 40 45
 Leu Glu Gly Ala Glu Arg Ala Val Arg Ala His Gly Ala Gly Trp
 50 55 60
 Asp Arg Gly Gly Tyr Arg Gly Arg Gly Ala Met Arg Arg Pro Ser
 65 70 75
 Val Arg Ala Ala Gly Leu Val Leu Cys Thr Leu Cys Tyr Leu Leu
 80 85 90
 Val Gly Ala Ala Val Phe Asp Ala Leu Glu Ser Glu Ala Glu Ser
 95 100 105
 Gly Arg Gln Arg Leu Leu Val Gln Lys Arg Gly Ala Leu Arg Arg
 110 115 120

Lys	Phe	Gly	Phe	Ser	Ala	Glu	Asp	Tyr	Arg	Glu	Leu	Glu	Arg	Leu			
				125											130		135
Ala	Leu	Gln	Ala	Glu	Pro	His	Arg	Ala	Gly	Arg	Gln	Trp	Lys	Phe			
				140											145		150
Pro	Gly	Ser	Phe	Tyr	Phe	Ala	Ile	Thr	Val	Ile	Thr	Thr	Ile	Glu			
				155											160		165
Tyr	Gly	His	Ala	Ala	Pro	Gly	Thr	Asp	Ser	Gly	Lys	Val	Phe	Cys			
				170											175		180
Met	Phe	Tyr	Ala	Leu	Leu	Gly	Ile	Pro	Leu	Thr	Leu	Val	Thr	Phe			
				185											190		195
Gln	Ser	Leu	Gly	Glu	Arg	Leu	Asn	Ala	Val	Val	Arg	Arg	Leu	Leu			
				200											205		210
Leu	Ala	Ala	Lys	Cys	Cys	Leu	Gly	Leu	Arg	Trp	Thr	Cys	Val	Ser			
				215											220		225
Thr	Glu	Asn	Leu	Val	Val	Ala	Gly	Leu	Leu	Ala	Cys	Ala	Ala	Thr			
				230											235		240
Leu	Ala	Leu	Gly	Ala	Val	Ala	Phe	Ser	His	Phe	Glu	Gly	Trp	Thr			
				245											250		255
Phe	Phe	His	Ala	Tyr	Tyr	Tyr	Cys	Phe	Ile	Thr	Leu	Thr	Thr	Ile			
				260											265		270
Gly	Phe	Gly	Asp	Phe	Val	Ala	Leu	Gln	Ser	Gly	Glu	Ala	Leu	Gln			
				275											280		285
Arg	Lys	Leu	Pro	Tyr	Val	Ala	Phe	Ser	Phe	Leu	Tyr	Ile	Leu	Leu			
				290											295		300
Gly	Leu	Thr	Val	Ile	Gly	Ala	Phe	Leu	Asn	Leu	Val	Val	Leu	Arg			
				305											310		315
Phe	Leu	Val	Ala	Ser	Ala	Asp	Trp	Pro	Glu	Arg	Ala	Ala	Arg	Thr			
				320											325		330
Pro	Ser	Pro	Arg	Pro	Pro	Gly	Ala	Pro	Glu	Ser	Arg	Gly	Leu	Trp			
				335											340		345
Leu	Pro	Arg	Arg	Pro	Ala	Arg	Ser	Val	Gly	Ser	Ala	Ser	Val	Phe			
				350											355		360
Cys	His	Val	His	Lys	Leu	Glu	Arg	Cys	Ala	Arg	Asp	Asn	Leu	Gly			
				365											370		375
Phe	Ser	Pro	Pro	Ser	Ser	Pro	Gly	Val	Val	Arg	Gly	Gly	Gln	Ala			
				380											385		390
Pro	Arg	Leu	Gly	Ala	Arg	Trp	Lys	Ser	Ile								
				395											400		

<210> 14

<211> 260

<212> PRT

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7474286CD1

<400> 14

Met	Met	Trp	Ser	Asn	Phe	Phe	Leu	Gln	Glu	Glu	Asn	Arg	Arg	Arg			
	1			5					10						15		
Gly	Ala	Ala	Gly	Arg	Arg	Arg	Ala	His	Gly	Gln	Gly	Arg	Ser	Gly			
				20					25						30		
Leu	Thr	Pro	Glu	Arg	Glu	Gly	Lys	Val	Lys	Leu	Ala	Leu	Leu	Leu			
				35					40						45		
Ala	Ala	Val	Gly	Ala	Thr	Leu	Ala	Val	Leu	Ser	Val	Gly	Thr	Glu			
				50					55						60		
Phe	Trp	Val	Glu	Leu	Asn	Thr	Tyr	Lys	Ala	Asn	Gly	Ser	Ala	Val			
				65					70						75		
Cys	Glu	Ala	Ala	His	Leu	Gly	Leu	Trp	Lys	Ala	Cys	Thr	Lys	Arg			
				80					85						90		
Leu	Trp	Gln	Ala	Asp	Val	Pro	Val	Asp	Arg	Asp	Thr	Cys	Gly	Pro			
				95					100						105		


```

Ala Glu Leu Pro Gly Glu Ala Asn Cys Thr Tyr Phe Lys Phe Phe
110 115 120
Thr Thr Gly Glu Asn Ala Arg Ile Phe Gln Arg Thr Thr Lys Lys
125 130 135
Glu Val Asn Leu Ala Ala Ala Val Ile Ala Val Leu Gly Leu Ala
140 145 150
Val Met Ala Leu Gly Cys Leu Cys Ile Ile Met Val Leu Ser Lys
155 160 165
Gly Ala Glu Phe Leu Leu Arg Val Gly Ala Val Cys Phe Gly Leu
170 175 180
Ser Gly Leu Leu Leu Leu Val Ser Leu Gly Val Phe Arg His Ser
185 190 195
Val Arg Ala Leu Leu Gln Arg Val Ser Pro Glu Pro Pro Pro Ala
200 205 210
Pro Arg Leu Thr Tyr Glu Tyr Ser Trp Ser Leu Gly Cys Gly Val
215 220 225
Gly Ala Gly Leu Ile Leu Leu Leu Gly Ala Gly Cys Phe Leu Leu
230 235 240
Leu Thr Leu Pro Ser Trp Pro Trp Gly Ser Leu Cys Pro Lys Arg
245 250 255
Gly His Arg Ala Thr
260

```

```

<210> 15
<211> 489
<212> PRT
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7472589CD1

```

```

<400> 15
Met Ser Ser Arg Ser Pro Arg Pro Pro Pro Arg Arg Ser Arg Arg
1 5 10 15
Arg Leu Pro Arg Pro Ser Cys Cys Cys Cys Cys Arg Arg Ser
20 25 30
His Leu Asn Glu Asp Thr Gly Arg Phe Val Leu Leu Ala Ala Leu
35 40 45
Ile Gly Leu Tyr Leu Val Ala Gly Ala Thr Val Phe Ser Ala Leu
50 55 60
Glu Ser Pro Gly Glu Ala Glu Ala Arg Ala Arg Trp Gly Ala Thr
65 70 75
Leu Arg Asn Phe Ser Ala Ala His Gly Val Ala Glu Pro Glu Leu
80 85 90
Arg Ala Phe Leu Arg His Tyr Glu Ala Ala Leu Ala Ala Gly Val
95 100 105
Arg Ala Asp Ala Leu Arg Pro Arg Trp Asp Phe Pro Gly Ala Phe
110 115 120
Tyr Phe Val Gly Thr Val Val Ser Thr Ile Val Arg Glu Glu Ser
125 130 135
Pro Pro Leu Ala Leu Thr Pro Gly Arg Leu Cys Ser Asn Thr Gly
140 145 150
Arg Leu Cys Asp Leu Thr Phe Lys Ser Tyr Ile Asn Ile Ala Lys
155 160 165
Glu Gln Glu His Pro Ala Ile Gln Gln Ser Phe Pro Arg Val Ser
170 175 180
Thr Val Ser Ser Glu Asn Arg Lys Glu Gly Phe Gly Met Thr Thr
185 190 195
Pro Ala Thr Val Gly Gly Lys Ala Phe Leu Ile Ala Tyr Gly Leu
200 205 210
Phe Gly Cys Ala Gly Thr Ile Leu Phe Phe Asn Leu Phe Leu Glu
215 220 225

```

Arg Ile Ile Ser Leu Leu Ala Phe Ile Met Arg Ala Cys Arg Glu
 230 235 240
 Arg Gln Leu Arg Arg Ser Gly Leu Leu Pro Ala Thr Phe Arg Arg
 245 250 255
 Gly Ser Ala Leu Ser Glu Ala Asp Ser Leu Ala Gly Trp Lys Pro
 260 265 270
 Ser Val Tyr His Val Leu Leu Ile Leu Gly Leu Phe Ala Val Leu
 275 280 285
 Leu Ser Cys Cys Ala Ser Ala Met Tyr Thr Ser Val Glu Gly Trp
 290 295 300
 Asp Tyr Val Asp Ser Leu Tyr Phe Cys Phe Val Thr Phe Ser Thr
 305 310 315
 Ile Gly Phe Gly Asp Leu Val Ser Ser Gln His Ala Ala Tyr Arg
 320 325 330
 Asn Gln Gly Leu Tyr Arg Leu Gly Asn Phe Leu Phe Ile Leu Leu
 335 340 345
 Gly Val Cys Cys Ile Tyr Ser Leu Phe Asn Val Ile Ser Ile Leu
 350 355 360
 Ile Lys Gln Val Leu Asn Trp Met Leu Arg Lys Leu Ser Cys Arg
 365 370 375
 Cys Cys Ala Arg Cys Cys Pro Ala Pro Gly Ala Pro Leu Ala Arg
 380 385 390
 Arg Asn Ala Ile Thr Pro Gly Ser Arg Leu Arg Arg Arg Leu Ala
 395 400 405
 Ala Leu Gly Ala Asp Pro Ala Ala Arg Asp Ser Asp Ala Glu Gly
 410 415 420
 Arg Arg Leu Ser Gly Glu Leu Ile Ser Met Arg Asp Leu Thr Ala
 425 430 435
 Ser Asn Lys Val Ser Leu Ala Leu Leu Gln Lys Gln Leu Ser Glu
 440 445 450
 Thr Ala Asn Gly Tyr Pro Arg Ser Val Cys Val Asn Thr Arg Gln
 455 460 465
 Asn Gly Phe Ser Gly Gly Val Gly Ala Leu Gly Ile Met Asn Asn
 470 475 480
 Arg Leu Ala Glu Thr Ser Ala Ser Arg
 485

<210> 16

<211> 1735

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 1784775CB1

<400> 16

atgtgctctcc ttgttctccc ccttgtcccc agggagtcag attacatcct acctgcagt 60
 cctggatggc gctccgaact tgcagcttcc ttctgcttt cctcttccc gctgctagac 120
 ctcttctccag ttgctttgcc accaggggca gcccaggag ccatagggtc agaggtgttg 180
 gcagggtgag tggcagctgt ggcttgatc agccacagc tggcctgtg ggtgttgcca 240
 cattccctcc atggccaact ccgggggtccc ttggccttgg ccctgttagc cttgctgcca 300
 gctccacccc tagtctgac cgtgttgttg cattggcagc gaggcacact tctgcccaca 360
 ctctccccag ggcacatggc ccgctctatg ttgtctatcc tgcagctggc tgcactcttg 420
 gcctatgcac tgggatgggc agctcctggg ggaccacgag aacctgggc tcaggagccc 480
 cctctgccg aggatcaaga acctgaggtg gctgaagatg gggagagttg gctgtcacgc 540
 ttttctctat cctggctggc accttctgct gccctgggg cctgtggaga cctccggcag 600
 cctcaggaca ttgcccgcct ccccacaga ctgcagccaa cctacctggc tctgtctctc 660
 caggcacact tggcaggagg ggcacggctg tggaggccct tgtatggggc ctttggacgg 720
 tgcctatctg cacttgagct gctgaagctg gtggggacca tgttgggatt ctcaggggcc 780
 ctgttctctt cctactgggt gggcttctgt gaagaggggc aggagccact aagccacggc 840
 ctgctctatg ctctgggggt agccgggtgg gctgtgctgg gtgctgtgct gcagaatcag 900
 tatgggtatg aggtatataa ggtaacactt caggcacggg gggctgtgct gaacatcctg 960

```

tactgcaagg ctttacagct ggggccagc cgccctccta ctggggaggc cctgaacct 1020
ctaggcaactg actctgaacg gctgcttaac ttgtctggga gcttccatga agcctggggc 1080
ctgcccctcgc aatctggccat caccctctac ctgctgtacc agcaggtagg cgtggccctc 1140
gtgggtggctc tcatcttggc actgctgctg gtaccctgtca caaagtgat tgcccaccgc 1200
atcatggcca gcaaccaggaa aatgctacag cacaaggatg cgcgggttaa gcttgtgaca 1260
gagctcgtga gtggcatctg ggtcatcaag ttctgcgggt gggagcagcg actcggagcg 1320
cgagttagag cctgccgggc tggagagctg gggcgactcc gggctatcaa atacctggat 1380
cgcgctctgt ttacctgtg ggtgcccta ccggttgtca tctccatctg tatcttcate 1440
acctatgtcc tcatggggca ccagctcact gccaccaagg tgaggaccag gaaggagggg 1500
gaccagcatc aaggagactt cagcgaaagt aagacagagc cttgggcccc ctgctctggc 1560
tgagaaggag ggagggatcc ctgactgcct catctctcaa ccaagggaaa actgaagaaa 1620
cctttttgtg ggggccttgg atataaacc ctcccctctg tgaaggagtt cctctcttcc 1680
ctctctctca cctttcactc cagcttctat tcagttccag gcttgggggt aggtc 1735

```

<210> 17

<211> 1041

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7473034CB1

<400> 17

```

atggttcaag catctgggca caggcggtcc acccggtgct ccaaatgggt ctctctggct 60
gtgatagcaa agatccaggaa aatatgggtg gaggaagatg agagggaagat ggtgcgagag 120
ttcttggccg agttcatgag cacatatgtc atgatgggat tcggccttgg ttctgtggcc 180
catatgggttc taataaaaaac atatgggagc tacccttggg tcaacttggg ttttggcttc 240
gggggtcacc a tgggagtcga cgtggcagcg cgcatctctg gagccacat gaatcagct 300
gtgaccttca ctacctgtgc gctgggcgcg gtgccctgga ggaagtctcc agtccatgtg 360
ctggggcagt tctggggctc ctctctggca gctgccacca tctacagtct ctctctcagg 420
gccattctcc acttttctgg tggagagctg atggtgacg ctcccttctg tacagctggc 480
atttttgcca cctaacttcc tgatcacatg acattgtggc ggggcttctc gaatgaggag 540
tggctgacca ggatgtctca gctgtgtctc ttcacatca cggaccagga gaacaacca 600
gcactgccag gaacacacgc gctggtgata agcatctcgt tggctcatcg cagggtgtcc 660
catggcataa acacaggata tgcacatcaat ccatacccc acccgcccc cagcatctc 720
acctctattg ctggctgggg caaacaggctc ttcagcgatg ggggaacctg gtggtgggtg 780
ccagtggtgg caccacttct ggggtgcctc ctagggtgga ccatctacct ggtcttctatt 840
ggctccacca tcccacggga gccctgaaa ttggaggact ctgtggcgta tgaagaccac 900
gggataacgg tattgcccaa gatgggatct catgaacca tgaatctctc cctcacctcc 960
atctccgtga gccctgccaa cagatctcca gtcacactct cccaccctct acatgaatcc 1020
atggccctag agcaactcta a

```

<210> 18

<211> 2367

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 1878581CB1

<400> 18

```

ggagcgctct cggggacgcg cgaggtcgcc gtggcgggag acgcgtttcc ggtggcggcg 60
ggagctgcac tgagcgggag ctgcgagcag cgcggggcgc agcccggggg aagcgtccgg 120
gacctgtctt ggagaaactac caccaaaacat taacatcaag gaacctcgat gggatcaaa 180
cactttctat ggcagagcca atcatttctt cactgttaac gaccacagga acattctgtt 240
aaccaacgaa caactcagga gtgcgagaaa aatagacat gattacagcg agggaaatt 300
tctctctggt ctacacagaaa atgaattgtg gagagcaag tacatctatg attcagctt 360
tcatctgcag actggtgaga agatgatatt gataggaaga atgtacgccc aggttccccat 420
gaacatgacc atcaacaggtt gtatgatgac gttttacag actacgcgg ctgtgctgt 480
ctggcagctg attaaacagt ccttcaatgc cgtcgtcaat tacaaccaaa gaagtggaga 540
cgcacccctc actgtcaatt agttgggaac agcttacggt tctgcaacaa ctgggtgcgt 600

```

```

agcaacagctg ctaggactca atgcattgac caagcatgtc tcaccactga taggacgttt 660
tgttcccttt gctgccttag ctgctgctaa ttgcattaat attccattaa tgaaggcaaa 720
ggaactccaa gtggccttc ccgtcaagg tgagaaatgg aaccgcttgg gggagctcgg 780
gaacgctgcg aaacaagcca tcacgcaagt tgcgtgtccc aggtattctc tggcagcccc 840
tggcctggcc atccctccat tcatatgaa cactttggaa aagaaagcct ttttgaagag 900
gtttcccatg atgagtgacac ccattcaagt tgggttagtt ggtctctggt tgggtctctc 960
tacacccctg tgttgtgccc gtttcctca gaaaagtccc atgtctgtga caagcttgga 1020
ggcgcaggtg caagctaaaga tccaagagag ccattcctgaa ttgcgacggc tgtactccaa 1080
taagggtatg taagcagagg aggaaacctc tgcagctcat tctgccactg caaagcttgt 1140
gtagccatgc tggtgagaaa aatctctgtc aacctgggtt ctccagatta cggaaacctt 1200
ttaaagatcc accattagcct tttagaataa acgtgctact ttaacagagc acctggcgtg 1260
ggccaaagtc ctgatactcc ctacactga atcatgttat gatttataga aatacctttc 1320
ctgtagcttt tatagtcatc gttttccaaa gacgatatac cagccctcac ccaggtttta 1380
aaaaagcact ggtaggcata gaataggctg tcatgtatgt gtcagtaaat gttctattga 1440
ttatcaatca gtgaaaaaag aatctgtttt aaatactga atttccatc cactcccat 1500
gcaaatcaag gagatctcag cagtgaactg ggaataataa aaagctctgg gctaattat 1560
aaaaaactac cctgaataat taagggcagt ttgctcttag ttgggggatt gcgcctggcc 1620
aatgaaggtg atgaagcttt tggatttggg gggtaaaaag tccctcacac ccctccaaa 1680
agtccagtc agacactcgc aacatgcctt cctgctgtga tcatatata catcagatt 1740
gtgagtgtat tgcttgggtt gacttttaat ttattgtttt ttgttcttat aaagatgata 1800
attctaccct cagttatttg actttatatt caattattta catcaataaa tcaactcat 1860
gaaatgtaca aatgtcaaat ttgggaagta tattcaata caatgctgta tgaagtggtg 1920
gaatccagtt catgtttttt tttttggtaa gaagtggagc tacagttcca gctacctaca 1980
tgtctttttt tgcctcctt atagatctct ttggcttcca gaaagatata gtgataatgt 2040
tgttatgaat cagtcaacaat gaattttact tgaattattg atgtgcatt ccaactcat 2100
tgaaaataat gaaacatgtt accactgttt acatcatctg tagtgatttc atagataata 2160
tatttaatat cagagattat gtttcaactc tgtagatgtt taacgtcata gacagtgtgc 2220
cctctgtatc cgtgagctct atatctgtga attcaacca gtttgatggg aaaaattttt 2280
tttttttttt ttttttgag acggagcttc gctctgtcac ccaggctgga gtgcagtgcg 2340
gtagtctcgg ctactgcaa gctccgc 2367

```

<210> 19

<211> 3343

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 2246292CB1

<220>

<221> unsure

<222> 57

<223> a, t, c, g, or other

<400> 19

```

cttgcagcgg cgcacgcggg atgggaggcg gggaggagca gcgggaagag cggacgncgc 60
accgctgcgc gcgcagctct caatgagcag cgcggaaact gcaccccaga ccgagcctgc 120
ctgcgcgcgc cctccacagag ctcaactggt gccaggtaac aggcctggcc tcgcctctgt 180
gatgatgatg gccttgcctc cgtgagctac aacctggcct ctacgacccc ccacactcca 240
accagcagga tgcggctgtg gaaggcgggt gtggtgact tggccttcac gagtgtggac 300
attctgcgtga ccacggccat ctatgtcttc agccacctgg accgcagcct cctggaggagc 360
attccgcact tcaaacattt tgactcgggt ctggtatctt gggcagcctg cctgtaccgc 420
agtgctcctc tgctggggag caccattggt gtggccaaga acagtgcgct ggggccccgg 480
cggctgcggg cctcgtggct ggtcatcagc ctctgtgccc tctctgtggg catcatcgcc 540
attgctgaagc tgctgctctt ctcaagggtg cgcaggccca tcogggaacc ctggttttgg 600
gcctgttctg tgtggacgta catttcactc ggcgcattct ctgctctgt gtgctgctgc 660
tccacgcgtc ggcaggcagc ccaggccctg gagccagggg cggccaccca ggcctgaggc 720
ttccctggga cggcgccggc acgcgccgag caggcgtctg gggccacgct cagaagctgc 780
ctctcttaca ccaagccgga cgtggccttc ctctgtgccc cctctcttct cctcatctgc 840
gcagctcgtg gagagacatt cctgccctac tacacggggc gcgcattga tgcagctgc 900
attccagaaa gcatggatca gttcagcagc gctgtcgtca ctgtgtgctc gctggccatt 960
ggcagctcat ttgcgcgagg tattcggggc ggcattttta cctcatatt tgcagcagct 1020

```

```

aacattcgcc ttcgaaactg tctcttcgct tcaactggtg cccaggagac aagcttcttt 1080
gagtagaacc gcacagaggga cctcatctcc cgctgacct cggacaaccac catggtcagc 1140
gacctgtgct cccagaaact caatgtcttc ctgcggaaca cagtcaaggt caccggcgctg 1200
gtggctttcca tggctcagcct ctcatggcag ctctctctgg tcaacctcat ggctctcccc 1260
atcatcatcga tgggtgtccaa catctacggc aagtactaca agaggctctc caaagaggtc 1320
cagaatgcgcc tggccagagc gagcaacacg gcggagagga ccatcagtcg catgaagact 1380
gtccggagct tcgccaatga ggaggaggag gcagagggtg acctgcccga gctgcagcag 1440
gtgtacaacg tgaaacaggaa ggaggcagct gcctacatgt actacgtctg ggcgacgggg 1500
tccgtgggct ccgtctacag tggcctgatg caggggatgg gggctgtcga gaaggtgttc 1560
gagtctcatcg accggcgagc gaccatggtg cagcatggca gcttggcccc cgacaacctc 1620
gaggggccggg tggactttga gaatgtgacc ttacacctac gcactcggcc ccacacccag 1680
gtccttcaga atgtctcctt cagcctgtcc cccggcaagg tgacggccct ggtggggccc 1740
tcgggcagtg ggaagagctc ctgtgtcaac atcctggaga acttctaccc cctggagggg 1800
ggccgggtgc tgtgtgacgg caagcccatc agcgctacg accacaagta cttgcaccgt 1860
gtgatctccc tggtagacca ggagcccggt ctgttcgccc gctccatcac ggataacatc 1920
tctatcgccc tgcgccactg gcctttcgag atggtggtgg agggccgaca gaaggccaat 1980
gcccacggct tcatcatgga actccaggac ggctacagca cagagacagg ggaggaaggg 2040
gcccactgtg caggtggcca gaagcagcgg tggccatgtg cccggctctc tgtgcggaac 2100
ccccagctcc tcatctcgtga tgaagccacc agcgctttgg atgccgagag cgagtatctg 2160
atccagcagg ccatccatgc caacctgcag aagcacacgg tactcatcat cgcgcacagg 2220
ctgagcaccc tggagcacgc gcacctcatt tgggtctcgg acaaggggcc gtagtgcagc 2280
cagggcaccc accagcagct gctggccagg ggccgctctc acgccaaagt ggtgcagcgg 2340
cagatgctgg ggcttcagcc gcgccagac ttcacagagt gccacaacga gccctgatgc 2400
aacgcgagtc acaaggcctg atggggggccc cctgctctc ccggtggggc agaggaacctg 2460
gtgccttgct ggcagatgtg cccacggagg ccccacagt cctccgagc ccaggcctgc 2520
agcactgaaa gacgacctgc catgtcccat ggatcacgc ttctgcate ttgccctgcg 2580
tcctcgcccc actccaggg cactccttac cctgctgccc ctgagccaac gccctcacgg 2640
acctccttag cctcctaagc aaaggtagag ctgccttttt aaactaggt tctaccaggg 2700
tttttaactg ttggtttgag gcaccccagt caactcctag atttcaaaaa cctttttcta 2760
atlgggagta atggcgggca ctttcaacca gatgtctcag aaactctcga gccaggagtg 2820
aatggccctt ccttagtagc ctgggggatg tcagagact aggcctctcc cctttaacct 2880
tcagagaag gggtctccct gtcccggagg gagaacgggt gaacgggatt ttccgtctct 2940
ccctcttgcc agctctgtga gtctggccag ggccggtagg gagcgtggag ggcatctgtc 3000
tgcatctgcc cgctgccaat ctaagccagt ctacatgtga accacacgaa acctcaactg 3060
ggggagtgccc gggctggcca ggtctggagg ggctcagagg gtgccccag cccggcaccc 3120
agcgctttcg cccctgctcc acccaacctt ggctggcagc ctccctcccc acaccccgcc 3180
ctgtgctctg ctgtctggag gccacgtgga tgttcatgag atgcattctc ttctgtcttt 3240
gggtgaatgg atggtggcaa agccacagat ctggctttgc cagaggtgtc aacatgttga 3300
gagaacccgg tcaataaagt gtactacctc ttaccctctaa aaa 3343

```

<210> 20

<211> 3517

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 5151730CB1

<400> 20

```

ccgacccgcg ctgggctggg ctgggctggg ctggggcggg cgcagggcgc agggcgggcg 60
gcgcggggga agacgcacgg ctcggctcgg ctctcccggg gagcgcccg ggactgcacc 120
gggacacggc cctctccgct tcgcgctgcc ctggccctcg ccccgggccc gggtagatga 180
gcgcgcggcc cgggggacat ggaagcgctg acgtgtggc ttctccccg gatattgccag 240
tgcgtctcgg tgcgggcaga ctccatcac cacactgggt ccatctcga ggagaaacgc 300
gccaaaggag acaggggtgt ccaagtggcg gtatccgac gttagcctcaa cgaatgacat 360
ctcagagagc agagaatcac ctactccatc aaggtcatcg agggcaacaa cccattccag 420
gctgtcagag aagcctgtga cctcatgacc caggggattt tggccttggt cagctccact 480
ggctgtgcct cctcaatgc ctgcagtcgc ccatgcacat gccacatgaa cccacagccc 540
tttgtccagg gcaaccgggg agggctgcga cgaacccgat gccacatgaa cccagccgct 600
gatgctgagg cctacacact ggcttcgaga ccaacctgc gcctcaatga tgcctatgat 660
aggctggtga cggagctcgc ctggcagaag ttcgctcatgt tctacgacag cgagtatgat 720
atccgtgggc ttcaaaagct tctggaccag gcctcgcggc tgggccttga cgtctcttta 780

```

```

caaaagggtgg acaagaacat tagccacgta ttcaccagcc tcttcaccac gatgaagaca 840
gaggagctga atcgctaccc ggacacgctt cgcgcgcgca tctgtctgct gagccacag 900
ggagcccact cctctatcaa cgaagccgtg gagaccaaac tggcttccaa ggacagccac 960
tgggtctctg tgaatgagga aatcagtgac ccgagatccc tggatctggc ccatgtgcc 1020
cttgggaagg taaccgtgtg ccggcacaatc ttctcgtctg caaaggacaa tcagaaatgc 1080
acgaggaaca accaccgcat cctctccctg ctctgcgacc cccaggaagg ctaccctcag 1140
atgtgctcaga tctccaacct ctatctgtat gacagtgttc tgatgtctgc caacgccttt 1200
acaggaagc ttggagaccg gaagtggcat agcatggcga gcttcagctc cataccgaaa 1260
tcactaacg catggaatcg tgggaagctcc atgcttgata ccatcaaaaa gggtccacatc 1320
actggctcga ctggggtgat tagtgagact ttgggcaagt caatgcctca tgtccagttc 1380
gaaactcctg gcactaccta tagtgagact ttgggcaagt caatgcctca tgtccagttc 1440
tggagctcag agaagggtct ggtgacgacg ttgcagagaa ggtcactggg cagccgcctc 1500
caagagttag ctctaaagg ggtgactgtc ttgggaagag ctctctgtat ggtggctgac 1560
aaactctcag gacagcccaa gcgtcacaaa gggtctccca tagatgtctc gtagctcact 1620
gccaaagctc tgggtcttaa atatgagatt taccaaagcc ctgatggcag gtacgtctac 1680
cagctccata acactcctg gaacgggatg atcggggagc tcacagcaa gagagcagac 1740
ttggccactc ctgctcacac catcaccca gagaggaga gcgtgtgga cttacgaag 1800
cgtgatctgg actattcagt ggggattcta attaaagaac ccgaggagaa aatcagatc 1860
ttctccctct tctgtccatt tgaattctgt gtgtggcctc gcattgcagc agccatccct 1920
tgtggttggt tctgtatatt tgtgtgaaac aggatacagg ctgtgagggc tcagatgtct 1980
gccaccccca ggcgctcagc ttctgcacat ctgcacagcg ccatctggat tgtctatgga 2040
gcctctgata agcaagggtg cgaatcttcc tgaactcca tggccatcgc catcgtgatg 2100
ggcagctggt ggcctctcac gctcattgtg tgcctcctcc acacagccaa cctgtctgcc 2160
ttctccacag tctccaggat ggacaacccc ataaggactt tccaggacct gtccaacaaa 2220
gtggaaatgt cttatggcac tgtccgggat tctgtctgat atgagtactt ctcgaaccaag 2280
ggccacaacc ccttgagaca ggacagcagc ttgtctgaac tctggcggac catcagcaa 2340
aacggyaggg ctgacaactg cgtgtccagt ccttcagaag gcatcaggaa ggcgaagaa 2400
gggaactacg ccttctctgt ggtgtggtcc gtggtggaat acgcatccct gacggatgac 2460
gactgtctcg tgaactgtac cggcaacagc atcagacga agggttacgg gattgcctcg 2520
cagcatgcca gccctacagc ggaactcttc tcccagagga tctgtgagct gcaggacaca 2580
ggggacactg atgtgctgaa gcagaagtg ttggccgaca tgggcccgtg tgaactcaac 2640
agccatgcca gcgccagcgc cgacggcaaa tccctcaagc tgcacagctt cgcgggggtc 2700
ttctgcatcc tggccattgg cctgctcctg gctgtcctg tggctgcctt ggaattgtgg 2760
tgaagacaga accgttgcca ccaggagacc cccaaggagg acaaaagaat gaacttggag 2820
caggtccacc ggcgcatgaa cagcctcatg gatgaagaca ttgctcacia gcagatttcc 2880
ccagctcgca ttagctcttc gccctggag atggggggcc tggctcccaa ccagacttgc 2940
gagccgacac gggagtacca gaacacccag ctctcgtgca gcaactttct ccagagcagc 3000
agcagccatg gcaccagccg gacactctca tcagggccca gcagcaactc gccgtgccc 3060
ctgagcagct cggcgaccat gccctccatg cagtgcaaac acaggtcacc caacgggggg 3120
ctgttccggc agagcccggt gaagaccccc atccccatgt ccttcagccc cgtgcctgca 3180
ggcgtctctc cagaggtctct ggacacctcc caggggacct ccatctgact gcgcgcctg 3240
ccctctcgcc caccctccca ccaccccgac cagcagagct tttataata agaaaaaac 3300
aacacaaacc acacacactc gcacacacac acatacacag agactcttct atttttctg 3360
tatcatatg taataaatga cagaatggag tggggtaaaa gtgtatttt atatttccca 3420
atttctgaag tcagtaaaaa aacacaaaaa ctgtatgaat gactttgtaa attttctct 3480
atatgaataa aaaggcaaat tactttgtaa aaaaaaa 3517

```

<210> 21

<211> 1248

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7472584CB1

<400> 21

```

atgaagaggc agaactgtgc gactctgtcc ctcatcgtct gcaccttcac ctacctgctg 60
gtgggcggcc cgtgttgcga cgcctctcag tcggaccacg agatgcgcga gaggagaaa 120
ctcaaaagccg aggagatccg gatcaagggg aagtaacaaa tcagcagcga ggactacagg 180
cagctggagcg tgggtatcct gcaigtggaa ccgcaaccgc ccgctgtcca gtggaattc 240
gccggtctct tctactttgc gatcacgctc atccacacca taggttatgg gcacgtgca 300
ctggcaccg atcggggcaa ggctctctgc atgttctacg ccgtgctggg catcccgctg 360

```

acactgtgtca	tgttccagag	cctggggcgag	cgcataaaca	ccttcgtgcy	ctacctgtct	420
aagcgcatata	agaagtgtct	tggcatgcgc	aacactgacg	tgtctatgga	gaacatggtg	480
actgtgggctt	tcttctcctg	catgggggacg	ctgtgcacatg	ggggcgccgcg	cttctccagc	540
tgtgaggagt	ggagcttctt	ccacgcctac	tactactgtt	tcatacagtt	gactaccatt	600
gggttcgggg	actactgtgc	cctgcagacc	aagggtgccc	tgacagaagaa	gcgcgtctac	660
gtggcctttta	gctttatgta	tatctgtgtg	gggctgcacg	tcactggggc	cttctcaaac	720
ctgtgtctcc	tcaggttctt	gaccatgaac	agtgaggatg	agcggcgggga	tgtctgaagag	780
agggcatccc	tcgcgggaaa	ccgcaacacg	atggtcattc	acatccctga	ggagccgcgg	840
ccagacggcg	ccaggtacaa	ggcggacgtc	ccggacctcg	agtcgtgtgt	ctctcgcacc	900
tgttaecgct	cgcaggacta	tggcgccgcg	tcggtggcac	cgcagaactc	cttctcgcgc	960
aagcttgccc	cccactactt	ccactccatc	tcttacaaga	tcgaggagat	ctcaccaagc	1020
acattaaaaa	acagcctctt	ccactcgcct	attagctcca	ttctctctgg	gttacacagc	1080
tttaccgacc	accagggctt	gatgaacacg	cggaaagtcg	tttaggggaa	ctaactgcac	1140
attcaagaga	ggcgctccgtg	gatgctgggt	ctcactgcca	aagccgaaca	cggcttcggg	1200
atttctgtcc	ttctcaagtg	gacctcttgc	tgtctgggc	ggaatgccc		1248

<210> 22

<211> 1770

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7472536CB1

<400> 22

atgtctgtcag	tcttcacccc	tggcctgttc	ctccccctcc	cagggccctc	cccggcctct	60
ctacataaag	ccgggggtac	tgggcctcag	gtcaggcccta	tggccatggc	cttcacagac	120
ctgtctggatg	ctctggggcag	catggggcgc	ttccagctca	accacacagc	cctgctgctg	180
ctgcctctgg	gcctctgggc	ctgccacaac	ttctctgcaga	acttcacgcg	cgtcttcccc	240
cccacacact	gcggggggcc	tgccaaccac	actgaggcct	ccaccaacga	ctcggggggc	300
tggcttaggg	ccaccatacc	cctggaccaa	cttggggccc	ctgagccctg	ccggcgcttc	360
accgaagcttc	agtgggcctc	gctgagcccc	aaactctcca	tcccggggcg	ggccacggag	420
ggcttgcaagg	acggctgggt	ctataaccgc	agtgttttcc	cgtccaccat	cgtgatggag	480
cagtgggatc	tgggtgtgtga	ggcccgcact	ctccagacac	tggcgacagt	cgtctacatt	540
gccgggggtgc	tgggtggggc	tgcctgtttt	ggcagcttgg	cagacaggct	gggctgcgaag	600
ggcccccctgg	tctgttctta	cctgcagctg	gcagcttcgg	gggcgcggcac	agcgtatttc	660
agctctctta	gtgcctattg	cgtcttccgg	ttcctgatgg	gcatgacctt	ctctggacat	720
attctcaact	ccgtctccct	gggtgagttg	atgccacac	ggggccggag	ctctggcagt	780
attttctggg	ggatattcct	caacctgggc	cagctcatcc	tggctggggg	agcctaccat	840
atttcgccct	ggcgttgctt	gcagtttgcc	atctctgtct	ctttctgtat	ctttttctct	900
tattcttggt	ggcttccaga	gtcatcccgc	tggctctctc	tgcattggcaa	gtcccagtta	960
gctgtacaga	atctgcagaa	gggtggctgca	atgaacggga	ggaagcagga	aggggaaagg	1020
gtgacaaag	cgggtgatgag	ctctcacatc	caaaagcga	gtcgaagtgt	ctgcaccttc	1080
aaactcaatc	tggacctctt	ccgaaccctgc	gccatccgca	aggtcacatg	ctgtccggcg	1140
ctgaggttat	ccaactctgt	ggcttactat	ggcctggcca	tggacttgca	gaagtgtggg	1200
ctcagcctat	acctgttgca	ggccctgttt	ggaatcatca	acatcccgcc	catgtctggt	1260
gccacgcaca	catgatatta	cgtggggcgc	cgtgccacgg	tggctctcct	cttcactctg	1320
gccggggctga	tgggtgatcgc	caacatgttt	gtgccagaag	gcacgcagat	cctgttgaca	1380
gccacgcaga	cgtggggcaa	aggctgcctg	gcagctctct	tcactgttgt	gtactctggt	1440
gccggcgagc	tgtaccccac	ggagatcagg	cagatgggga	tgggttttgc	ctctgtccac	1500
cctgcctctg	ggggcctgac	ggcgcccttg	gttaccacac	ttgggggaa	cagcaccatt	1560
ctgcacccgc	tgaactttgg	ggccaccgca	atcttggtgt	ggctggccgt	ctgcttctct	1620
actgagaccc	gcacacatgcc	cctgggtggag	accatcgacg	ccatggatga	ggaggtcaaa	1680
gaaggtctctt	cgaagaacaa	tgtagaagag	aagagtgaag	aaatttctct	tcagcagctg	1740
agagcatctc	ccctcaaaa	gacctatcaa				1770

<210> 23

<211> 2544

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7473422CB1

<400> 23

atgacgatcc gccgcgaacc gctgatgctg accttgccgc gcgcgcgtgt gagccctggc 60
 atcgccgcctc cgccgcgcctt ggcagacgca ccggtcacc tgaccatgta caacggctcag 120
 cacaaagaaa tcggcggaagc catcgccaag gcctacgagg caaagaccgg catcccaactc 180
 gatataccga agggcagcag caaccagctg gccagccaga tcatcgagga gggcgaccgt 240
 tgcgccagccg cctgcatctc caccgaagag tccccgccg tgaacaacct gggcggaactg 300
 ggccctgctgg cgaagatcga cgagccacc gcgaacatgt tgcaccaaga gtatgtggcg 360
 gccaaaggcca cctggatggg catcaccgcg cgacgcgca tcgtgtgtgt caacccgaaag 420
 aaggctgatg aaaaagacct gccgaccaca gtgatggact tcgcccaacc tgatgtggaa 480
 ggccgcgtcg cctagctacc caccagcggt gcattccagg agcagcgctg ggccatcctg 540
 aagatgcatt gtcgtgaagc caccgaagaa tggctgacc gccttaaacg cttcgccaaa 600
 acatacacca acaacatggt gcacctcaaa gccgtggaaa aaggtagagt ggctgcggtg 660
 ctggtgaaca actactactg gtacgcactt gaacgcgaac gcggcgaagt cgacaccaag 720
 ctctactacc ttggcagatgg cgatgccggc aacctggtag ccatctctgg ccgcgctgtg 780
 tccaagccca gcaagcaccg gaaagaagcc caggcactcg tcaactggat ggccagcgaa 840
 gaaaggccaac gtgtgatcac ccagaccacc gccgagtacc cgctgcacaa gggcatggtt 900
 tcgcagccgg gccctaagcc gttcgaagac ctgcgcccg gcgaaatctc gccagccgac 960
 ctggggcaatg ccgaggaagc catcgagctt gaacgcgagc gcggcccggt acgcttcgta 1020
 ccgcgcgcga agcgcccttc catctgggtg gtgctgcctg tgctgttctt ggtggcgatg 1080
 agcttgtctc cgctgctgta tgtgccatc aaagcctggg aagccggctg gcgtgaagcc 1140
 tgcacactgc tgtggcgccc ctltgtctgg gggctgatcg gcaataacct gatgctgatg 1200
 ctcggggtga cattggcctg catggtggct gggctggccc tggcctggct gctggagccg 1260
 agcaacactg ctcggccgcg gctgtggggc gtgggtgttt gcctgcctt cgtgtgccg 1320
 tcttctctga cgagtttacc ctgggtgtcg ctgagctcgg acttcgaagg cctgcgctgg 1380
 gccatctcgg tcatggcctt gtccaagtac ccatttgtgt tcctgcgggt ggccgcacc 1440
 ctgcgcgaac ctcgacacct gctggaggag tcggcgcgca ccctgggttg tagccgctgg 1500
 ggcgtgttca tcaaggctac cttgcccgct ctgtggccct cgatgctcgg cggagcgctg 1560
 ctgatcgccc tgcacatgct ggtggagttc ggtgcgttgt cgactctcgg cctgcagacc 1620
 ttcaccacgg cgatctacca gcagttcgaa ctggaattca gcaatgccaa ccgcccgatg 1680
 ctgtcttgct tgctgttgcc gatgtgcctg gtgatgctgt ggctggaatt gcgtgtacgt 1740
 ggcaagagcc gccatgtgag catcgccagc ggcgtggcac gccgcgcga acccgtagca 1800
 ctgcgttgct gggccgtacc ggccgaagta ctctgcgtgt ccctggcagt gctggcgagc 1860
 ggtatccacc tggccatgct cggctactgg ctgagcgtgt gttcgtcggc agccttcccg 1920
 tggggagcca tctccaagcc gctgttcacc tcgctgtcgg tgcgtctgg ccgtgccggt 1980
 ttctgtgtgc tctggcgct accgataagc ttctgtgtgt tgcgtacaa aggcctctct 2040
 gcgatctggg ccgagcgctt gccgtacctg ctgcacgccc tgcccgcct ggtgattgca 2100
 ctgacacttg tttctctcgc cctgcactac gtgcgcgcgc gtaccagac caggcgcttg 2160
 ctgctcttgg gctatgcgct gctgttctcg ccattggcgc agtcaccggt gcgaccgcg 2220
 ctgaacaagg cctcgccaac actggaggaa gccgcgcgca ccttgggtgc cagcagcttc 2280
 acggcattct gccgggtgac cctgccgata atcttcccgg ccattggcgc agcatlttgcg 2340
 ctggtgtttc tggatgccat gaaagaactg acagctaccc tgcgtcgcag ccgcaccggg 2400
 atgaccaccc ttgctaccga ggtgtgggag catacggcca acgtcgagtt ccgcggcgcg 2460
 cggccctatg cagccttctg gatcgtgggt tcaggcctgc cggtttatct gctgaccacc 2520
 cggatgtacc tgaacaagcg ataa 2544

<210> 24

<211> 2871

<212> DNA

<213> Homo sapiens

<230>

<221> misc_feature

<223> Incyte ID No: 3864715CB1

<220>

<221> unsure

<222> 2847

<223> a, t, c, g, or other

<400> 24


```

gcttctgccc  tgcctgctgt  gtccggagcc  gtccagcgac  caccatgggtg  aggcctcgctc  60
tgcccaaccc  cggcctagac  gcccggtacc  cgctccctggc  tgagctggag  accatcgagc  120
aggaggaggg  cagctccccc  ccgaagtggg  acaacaaggg  gcagtacatg  ctacacctgcc  180
tgaggctctg  cgtgggcctc  ggcaacgtgt  ggcgcttccc  ctacctgtgt  cagagccacg  240
gaggaggagc  ctctcatgat  ccgttctctc  tcttgcctggt  cctggagggg  atccccctgc  300
tgtacctgga  gttcgccatc  gggcagcgcc  tgcggcgggg  cagcctgggt  gtgtggagct  360
ccatccaccc  ggccctgaag  ggccctaggg  ttggcctccat  gctcacgtcc  tctatggttg  420
gactgtatta  caacaccatc  atctcctgga  ccatgtggta  cttattcaac  tctctccagg  480
agcctctgcc  ctggagcgac  ttggccgtca  acgagaacca  gacagggtat  gtggacaggt  540
gcgccaggag  ctccctctgt  gactacttct  ggtaccgaga  gacgtcaca  atctccact  600
ccatcagcga  ctgggctccc  atccagtgtt  ggaigtctct  gtgcttgccc  tggcgtatga  660
cgctctgtta  catgtgcacc  atccgcggca  tgcgaccacc  cgggaaggcg  gtgtacatca  720
ctccacgctg  gccctatgtc  gtccctgacca  tcttctctcat  ccgaggcctg  acgtgaaag  780
gcgcacccaa  tggcatcgct  ttctctctca  cgcccaacct  cagcgagctg  cccgacccg  840
acactggct  ggcgcggggc  gcacaggctc  tcttctctct  ctccctggcc  tccggggccc  900
tcatctctct  ctccagctac  aactctgtgc  acaacaacct  cgagaaggac  tgggtgatgt  960
tgtccatcat  caacgcccct  acatcggtgt  atgtggccat  cgtggtctac  tccgtcatgt  1020
ggttccgcgc  cagcgagcgc  taagcagact  gcttcagcac  gaacatccct  accctcatca  1080
acgggtctga  cctgcctgaa  ggcaacgtga  cccaggagaa  ctttgtggac  atgcagcagc  1140
ggtgcaacgc  ctccgacccc  gcggcctacg  cgcagctggt  ttccagacc  tgcgacatca  1200
agcctctctc  ctacagagcc  gtggagggca  caggcctggc  ctctcatcgt  tccacaggag  1260
ccatcaccaa  gatgcgctgt  tcccactgtt  ggtctgtctc  ctctctcatt  atgctctctc  1320
gctcggggct  gtcctctatg  ttgggaaca  tggagggcgt  cgttgtgccc  ctgcaggacc  1380
tcagagtcat  gcccccgaag  tggcccaagg  aggtgctcac  aggcctcatc  tgcctgggga  1440
catctctcat  tggcttcatc  ttacgcctga  actccggcca  gtactggctc  tctctgtggt  1500
acagctatgc  cgtgctcatc  cccctgctca  tcatcgctca  ctgcgagatg  ttctctgtgt  1560
tctacgtgta  cggtgtggac  aggttcaata  aggacatcga  gttcatgatc  ggccacaagc  1620
ccaactctct  ctggcaagtc  acgtggcgcg  ttgtcagccc  cctgctcatg  ctgatcatct  1680
tctctctctt  ctctgtggtg  gaggtcagtc  aggagctcgc  ctacagcatc  tgggacctgt  1740
gtctcaggga  atttcccaaa  tcccagaaga  tctctactcc  gaactggggt  tatgtgtgtg  1800
tgggtatgtg  ggctggagtg  cctccctcca  ccatccctgg  ctatgccact  tacaagctca  1860
tcaggaaaca  ctgcgagaag  ccaggggacc  atcaggggct  ggtgagcaca  tctctccacc  1920
ctcccatgaa  cggggacctg  aagtactgag  aaggcccatc  ccacggcgct  ccatacactg  1980
gtgtcaggga  aggaggaacc  agcaagacct  gtgggtgtgg  ggcgggctg  cactgcatgt  2040
tgtgttaagc  tgagtgtatg  ctctgtgtgt  agtgtgtgtc  ttgtacacgc  atgtgccatg  2100
tgtgcagata  tgtatctgtg  gtgcatgtac  atgcatgggc  actgtgagtg  tgcacgtgta  2160
tgcaacata  tacatgtgtg  tgggtgtgtg  tatgtatgtg  gcatgtgcc  tgtgtgcaga  2220
tgtgtcatgt  catgtacatg  catgtacatg  tatgtacatg  tatgtacatg  gcaagtgctg  2280
atgcataaac  atgtgtgcga  tatttgtctg  ccgtgtgtgt  gcatgtatat  atagacatat  2340
gtgtcatgt  tgtgtgtggt  gtgcatagt  gtgcatagt  acgtgtatgc  atgcactgac  2400
atgtgctcgt  acaatggggt  tccacatgca  cgtgtatatg  tatatctgtg  agtgtatata  2460
cagcatgca  attgtgtgta  tgtgtgttct  gtgtgtcgt  ttgcaagtat  atatgcacat  2520
tgtatatgt  acatgtatgc  ctgtgtgacg  ttgtgtatgt  tgagcatgtg  tacgtgtgtg  2580
tatacgtgtg  tgtgttatat  gtgtgtgtct  gtagtatgtg  tgtgtatgtg  gtgtgtatgt  2640
tgctcgtgtg  tgtgcataat  caggcaggtg  tgcatttgtg  catccccagt  tgtatgtatg  2700
tgcgcataat  gacacgcagt  gacacgcata  tggacacata  tggacacata  tatggacagt  2760
tgtgtgatgt  tgtgtgtaca  cgtcgtctgg  acacatgctc  gccactcggg  cccagcagtg  2820
ccctctgtgt  ttggggatcc  actattntaa  gccgcaccac  cgcgtgactcc  a  2871

```

```

<210> 25
<211> 2141
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 1734724CB1

```

```

<400> 25
gcggcgaccg  cgggacggcg  agaggcacgc  ggcgggaggg  gaccggaate  cgcagctccg  60
gcgcggccat  ggcggccaac  gacaacgtga  ccttgcctct  cgccccctgt  ctgcgggaca  120
actaacccct  gcggcccaat  gccagcagcc  ttggcccccg  cagcaacctc  gccctcgcce  180
ctgectccag  cgcgcggccc  ggccctgggc  tcagcctcgg  gccgtatccg  agcttcggct  240

```

```

tcagccccg cccactccg accccggagc ccacgaccag cggcctcgg ggcggcgagg 300
cgagccacgg cccctcccc gttccctcgg cctcgggcgc ccacgcgcgt cccgttctgg 360
gacaccccg tgaccacgg gctgaacgtg tctcgggcgc ccgcctgtgc atcaccatgc 420
tgggcctggg ctgcacgggt gacgtgaacc actcgggggc gacgtccgt cgcccgctgg 480
cgggcgtgct ggcagctctg ccagttcggc ctctcgccgc tctgtggcct cctgtgggac 540
ctgcctctca agctggagca ggtggccgcg gtgggtgct cctgtgtggc tgccttcccg 600
cgggcaatct ctcaatctt atgtccctgc tggttgacg gcacatgaac ctacagacct 660
ctgctctctt ggcactctcc tcggatgtag gttctgccca gacttcaacc ccgggactgt 720
cagctctccc gttccacct tactcaacat acaagaaaaa ggttagctgg ctggttgact 780
caaaagctgt tctgatttct gcacattccc tttcttgcag catcatcatg accactctct 840
ccacgcttct ggcctcgctg ttgatgcgcc tggcctgtg gatctacagc tgggcttgga 900
tcaacacccc tatcgtgcag ttaactaccc tagggaccgt gacctgact ctgcgcaga 960
ctctctaac ttcgggttg ggcgtcttca ttgcctaca atacagcgg gtggttgact 1020
acattgtgaa ggtttccctg tggctctctg tagtgactct ggtgtccct ttcataatga 1080
cgggcaatat ctgaggacct gaactcgtgg caagtatccc tgcagctgtt tatgtgatag 1140
caatttttat gcctttggca ggcacgtctt caggttatgt ttttagtact ctcttccatc 1200
ttccacccaa ctgcaagagg actgtatgtc tggaaacagg tagtcagaat gtgcagctct 1260
gtacagccat tctaaaactg gcccttccac cgcaattcat aggaagcatg tcatgttttc 1320
ctgtcttata tgcacttttc cagtctgcag aagcggggat tttgtttta atctataaaa 1380
tgtagtgaag tgaatgttg cacaacggag atctcttaga tgaagatgaa gatacagata 1440
ttctctataa aaaactaaaa gaagaggaaa tggcagacac tgcactggc acagtataaa 1500
cagaaaatat aataatgatg gaaaccctct agacttctct ctaaatgtgg agatacacag 1560
gagcttctat ctgtgtgaaa tattgttcca tatttatggc ctgtgtgtag tgcacgtgtc 1620
aaacataaaa ataacactgg ttacatcatc acatgaaca attctgatct ttttaagggt 1680
cactgtgtga ttaaccaaac gttgtcacaa attacaaatc aatgctgtaa tataatttgc 1740
acctggaatg gtaacagtga agcctgaatt aaatgtggtt ttttagtttt accatcacca 1800
attttctaga ctgttgcaaa tacagaatct attagaaaa agggctctgg aatgtgtaga 1860
tttttgcgca ctatgaggaa aaacaagcta tctttgtaaa ccaataatga gtttaagtta 1920
attgttataa aaaaaaaagt gtgcttgctc tacttaaaa tctcacaa gtgtaatttt 1980
gacctgtatt cagaagaatt ccaaaacagg tcagttaaat aaggaatat agtatttgtc 2040
aaaccagtat cagagaaaa ttacattaat gtatttgtt actgtatctg gtatctactt 2100
attaatgaat aatcaacatt tttctagtga aaaaaaaaa a 2141

```

<210> 26

<211> 1902

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 1563237CB1

<400> 26

```

ccgggtagtg agcggaggga caggaagggt agggcaagaa agggagagg gacaggagg 60
aagggtgggc caaagcgggt aaaaaggagg gccagccagt tctgtgggg agagtggcg 120
agccggcggg cgaggagtgc agggctctga gccggggaga ggagaggaga gaagagccgc 180
ggggggccca gccggagcc aggatgcgcc gcgcgcgcgc ccgggagcag ccccgctgc 240
ccggggagcg ccagcccgct ctgcctcgcg gtgcgcgggg cctcgacagc tggcgggcgg 300
cgggcgggcg ggcctgtgct ctggtggaga tctgtggagc gcgcgccttc tccggcgcta 360
cgcccaacct cgtgctgtac ctcaacagca ccaactcaa ctggaccggc gagcaggcga 420
cgccgcgcgc cgtggtatct ctggggcgct cctacctgtc gcgcgcctg ggcgctggc 480
tggcgcgagt gtacctgggc cgtaccgcgc cggctcgctc cagcctgctg ctctacctgt 540
cgcctcggg cctgctgccc gccacccgct tccccgacg cgcagctcc tctcgggag 600
agatgccgcg ctgcgcgctg ggaacctgct gccctcgcg cggctgccc cgcctcgcc 660
ccagccctca ctgcgcgccc gtctctacg cggcctgct gctactcgcc ctggcgccga 720
cgtcgtctcg gagcaacctc acctctctg tgcggaccca ggtgatgga ctggcgccg 780
acgccaccgc ccgctctctc aactggtttt actggagcct caacctgggt gctgtgctgt 840
cgtcgtggtt ggtggcgctt attcagcaga acatcagct cctcgctggc tacagacacc 900
ctgtgggctg tgtggcgctg gcatttttca tctctctct tccaccccc gctctcatca 960
cacaagcccc gatggcgcag caagtgtcct ctatgcttaa gctcgtctc caaaactgct 1020
gcccccagct gtggcaacga cactcgccca gagaccgta atgtgccgc gtgcgtggcg 1080
acgagaggtc tccccagcca ggggcttccc cgaagagga catcgccaac tcccaggtgc 1140
tgggtgaagat ctgcgcccgc atgtgtgacc tgggtgccca ctggatggtc tacttccaga 1200

```

tgcagtcac	ctatgtccctg	cagggtcttc	acctccacat	cccaaacatt	ttcccagcca	1260
accgggccaa	cactctctgtg	gccctcgagag	cccagggcag	cagctacacg	gagctctgga	1320
gatggagcgc	cttacacttac	atccaccaca	acgagaccgt	gtcccagcag	atgtggggagg	1380
tcctgtacaa	cgggccacca	ctgtccatct	ggtggcagat	cctcagtagt	ctgtccatgt	1440
ggatcagtaga	gatctttggc	agcatcccgag	gccctggagtt	tgcttactca	gaggccccgc	1500
gtctccatgca	ggggcccatc	atgggcattct	tcctctgcct	gtcgggggtg	gggtcactgt	1560
tgggctccag	ccatgtggca	ctgtctgtcct	tgcccggggg	ctggctgcac	tgcccacaagg	1620
attttgggaa	ctacaacaat	tgccggatgg	acctctactt	cttctctgct	ggtccatctt	1680
aggccgtcac	ggctctccta	tttgtctgga	tcgtggagcg	ctatgagagg	gcgtccacag	1740
gcccagcctc	ccacagcgtg	ttcagcaggg	acaggggtg	aacaggccct	attccagccc	1800
ccttctgtca	ctctaccgga	cagacggcag	cagctccacg	ctctggtttc	ttctcggttt	1860
attctgttag	aatgaatgtg	ttccataaat	aagggcagtg	tc		1902

<210> 27

<211> 4125

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7473443CB1

<400> 27

atgggggaaa	aacagtgcac	aaaggctaaa	aattccaaaa	accagaatgc	ctctctctct	60
caaaaggatc	acaactcctc	gccagcaggg	gaacaaaact	ggatggagaa	tgaattgaca	120
cgagcaggct	tcagaagggtg	ggtggttaata	aactcctgca	agctaaggga	cgatgtttta	180
accctaagtta	aggaagccaa	gaaccttgaa	aaaagggttag	gcgaattgct	aactagaata	240
accagtttag	aggaagaacat	aaatgacctg	atggagctga	aaacacacg	accgagaact	300
ctgtatgtcat	acataagtat	cagttagccga	attgatcaag	cagaaaaaag	gatctacagag	360
ctatgaagatc	aacttaatga	aataaagcgt	gaagacaaga	ttagagaaaa	aaatgaaaaag	420
gatgaacaag	gcctccaaga	aatatgggac	tatgtgaaaa	gacaaaacct	acatttgatt	480
gggttacctg	gcctgctcta	ttcagacatg	tgccgcctcc	tcctctctcc	aagaaatcag	540
ctctgcttgc	aggctttgga	acgtggagtc	atcttggagg	tcaaatcgct	cgtttgcagc	600
acgcaggctg	gggcagcgcg	gcgtggtgta	aagataaagt	tcaaggggaa	aggattttct	660
gtgggtgtcag	tcgtaggaac	cttgcaatgg	ttgtcttggg	ccagagccgc	gcattctcca	720
cactggcgct	ttctacgttg	gatggcagcc	ctgtggagtg	tgccagccaa	aactggccct	780
ttcccacata	gcctcacggg	tcagagaggg	aaccggggct	cggagtctct	ctctatccct	840
cgaggcgtgc	ccaaggattt	cagcacggga	acatcacgcc	aactaaggcg	agccatgggg	900
ctggcccctg	aggggcgagg	tttccaggcc	ttcttcccca	ggcccaccat	gcttgcacat	960
cccaactctc	tcgaaaacct	cagcttccag	agccgctgga	ttcccctcca	gccaatggcc	1020
gtgggctggg	cttttgtgca	gaagacctcg	gccctctcgt	ggctgctgct	tcctatggacc	1080
ttccctgtccc	ctcgctgggg	acaggccaag	attcctctgg	aaacagtga	gcatgtggct	1140
cgacaacttcg	cgaggggacct	gtataaacat	gtgaccaaat	actcaggctc	tcctctgtctg	1200
cagaagaagt	acaaggatgt	ggagttccag	ctgaagatcg	aggaggtgga	tgctttggag	1260
ctgtgtgagga	agttcttcaga	ggacatggag	aaactgctgc	ggaggaaagt	cgaggccgctg	1320
cagaacttgg	gggaagctgc	cgaggagggc	gacctgaacc	acgaattcaa	tgaattccctg	1380
gtggaaacttg	gcgtgggaat	tgccgtgggg	atgtccgtga	cgcagtcagg	cggtggagatt	1440
ggcgtggggga	gttccctggac	gcgtcccgcc	gtgggagttg	gcgtggggat	gtccataaacg	1500
ctgtccggcg	tgggagttgg	cgctggggatg	tcctgtgagg	agtcggggct	gggagattggc	1560
ctgggaatgt	cgtgtacgca	gtccggcggt	ggagttggcg	tggggaatgc	cgctgacgcag	1620
ctcgccgtgg	gagttggcgt	gggggaigtcc	gtgaggacat	ccggcgctggg	agttggcgtg	1680
tgcaagtctcg	tgaccgcagt	ctgggggggtg	ctcagtgccc	agcgcgcggc	cgcggtgtgt	1740
tggttagaact	ctggcgcgtg	ccggcccgccg	gccctctcgt	ctcttaccct	gcgcgctttc	1800
ttctactctc	ttccgcctgt	ttccgcctgt	ccggggccctg	ccgtggggct	caccgcgttc	1860
ccccagttcg	actattacac	ctcgggtctg	atcaacagaa	gggacagaaa	ggggcaacttc	1920
gtggagctgg	ggcccgagtt	ctctctctgg	ctccaatggc	acttcagcaa	cttgccgggtg	1980
aaacactcca	tcagcagcgt	gcagctggcc	accacagctg	acacaaagaa	cccgagattt	2040
ttaaattggag	ttcacatgtc	tgaaagcttg	aatgtgtctt	tcgtggagaa	cttccagaga	2100
gaccaacagt	tgacctggca	atatattggc	agtgcaactg	gattcttcag	gactatccac	2160
ggtataaaat	ggcacacttc	tgagaaatga	gtcataactt	ttgactgcgg	accgcggccc	2220
tggtacaact	aaagtctcag	ttctcccgag	gacatagttc	tttgggtgga	cgtagagcgg	2280
agtattgaagg	ggctgaggat	gactattggc	aagcacacca	tcaccacact	cttgacacac	2340
ctggggggaga	atgaactcat	taatatcata	gcgtacaatg	actacgtcca	ttacatcgag	2400

```

ccttggtttta aagggatcct cgtccaggcg gaccgagaca atcgagagca ttccaactg 2460
ctggtaggag agtgatgggt caaagggtgt ggggtcgtgg accaagccct gagagaagcc 2520
ttccagatcc tgaagcagtt ccaagagggc aagcaaggaa gcctctgcaa ccaggccatc 2580
atgctcatca cgcagcggcg cgtggaggac tacgaagccg tgtttgagaa gtataactgg 2640
ccagactgtca aggtcccgat ttctacttac ctcatgtgga gagaagtgtc tttgtctgac 2700
cgcatgaagt ggatgtcacg caacaacaaa ggctactaca cgcagatctc aacgtcgggc 2760
gacaccaggg agaacgtgat ggaataacct cacgtgtctca gccgcccatc ggtcatcaac 2820
cacgacacag acatcatctg gacagaggcc tacatggaca gcaagctcct cagctcgcag 2880
ctcgagagcc tgacactgct caccactgtg gccatgcagc tcttcagcaa gaagaaagaa 2940
acgcgatccc atggcattct cctgggtgtg gtggggctca agtggacctc gagagagctg 3000
ctgaagctgg cgcgccggta aaagcttggg gtgcacggat acgctcttct gaacaccaac 3060
aatggctaca tctctcccca tcccgcacct cgsgccctgt acagagaggg gaagaactca 3120
aaaccacaa ctaactacaa cagtgtggat ctctcggagc ttgagtggga agaccagctt 3180
gaatctctga gaacagccat gatcaatagg gaaacagcta ctctctgat ggaatgtga 3240
gttcccgatg ataaaggaaa gcgagttctt ttctgacca atgactactt ctccacggac 3300
atcacgcaca cccctttcag ttgtgggggt gtgctgtccc ggggcacagg agaatacatc 3360
ctttctggga acagctctgt ggaagaaggc ctgcattgact gctctcaacc aacttgccc 3420
ctggccgggt atggtatcta ctgcatacaa gatattgacc cagacacacc gaagctcagc 3480
cagctagagg ccatgtccg ctctctcacc aggaaggacc cagacttgga gtgtgacgag 3540
gagctgtctc gggaggtgct gtttgacgcg gtgggtacag ccccatcgga agcttactgg 3600
acagcccgct cctcaacat gtcccaggag tctgaacctg tggtagacat gccctctctg 3660
ggcaccggcg ctggccctct gagaagcagc ttgtctgtgt gctccgagaa ggtctccgac 3720
aggaagttcc ttgacctga ggacgaggcc agcgtgttca ccttgagccg ctctccctgt 3780
tgttaccgcc aggcctcaga gcatcctgct ggcagctctc tcttcaacct ccgctgggca 3840
gaaggaccag gacgcccctc tgccaaaggc ctctcacacc ctcttgcca aacctcctc 3900
aagcgtctgt atggaaaaaa gtccctggagc tgatgctgtg ggaggccagc gatcccgggc 3960
atctctctgt gaccaggaca gggaaggcaa acagcaggga ggagctcac ccgcccctc 4020
cgggaatccc gccctctcta cctggtgcta tgcccctact cccactctgt ctgttccctc 4080
cacacagctg caataacta ctatgtctta caaacaataa aaaaa 4125

```

<210> 28

<211> 2460

<212> DNA

<213> Homo sapiens

<220>

<221> misc_feature

<223> Incyte ID No: 7473438CB1

<400> 28

```

gggagggagg gatgggaaa ggagtgaag gaaggaagga aggaaggagg ggaaggagg 60
aggaaggcag gggaccgggg aggggaggga gggtaaaagg aaggaaggaa ggaaggagg 120
aaggacggaa gactggcacg aggtcacaca atgagttagt ggcgggaaca ggaactctgc 180
agggccgggg ttgggacct ttctctagat ctactatct acagacacat gaaccocggg 240
caggcgagtg ggagggggac aggggagcgg ttctctccgc cccacccgtt ggcgatccca 300
gcctccaggt tcccgcgagt cggcccgcct cggccgtcac aaccttgccc cgtggggccc 360
gggttggagg gggcgagagc cgcggtccgg gcacacggag caggttggga cgcggcgagg 420
taccggggcc gggcgcccat gcgagagccc agcgtgcgag cggcggggct ggtctctgtc 480
acctctgttt acctgctggt gggcgctgct gtcttcgagc cgctcgagtc cgaagcggaa 540
agcgccgcgc agcgactgct ggtccagaag cggggcgctc tccggaggaa gtctggcttc 600
tcggccgagg actaccgga gctggagcgc ctggcgctcc aggtcgagcc caacgcgccc 660
ggcccgcggt ggaagttccc cggctctctc tacttgcca tcaacctcat cactacatca 720
gagtacggcc acccgccgccc ggttacggac tccggcaagg tctctgcat gctctacgag 780
ctctctggga tcccgctgac gctggtcact ttccagagcc tgggcgaagg gactgagcgc 840
gtggtgcgag gctctctggt ggcgggcaag tgctgctctc aggtcgagcc cgcgtgctgt 900
tccacggaga acctgggtgt ggcgggggct ctggcggtgt ccgcacacct ggcctctggg 960
gcctgtgctct tctcgactt cgaggcgctg acctctctcc acgctacta ctaactgctc 1020
atcacctcaca ccacatcgg ctctcgagac ttctgtgac tgcagagcgg cgactgctgt 1080
cagaggaagc tctccactgt ggccttcagc ttctctaca tctctctggg gctcacggtc 1140
attggcgctct tcttcaacct ggtggtcctg cgtctctcgc ttgccagcgc cgaactggcc 1200
gagcgcgctc cgcgaccccc cagcccgccg ccccgggggg cgcgtggcctc cgttggtctc 1260
tggtctcccc gcgcggccgc ccgctccgtg ggtctcgctc ctgtctctgt ccaagtgcac 1320
aagctggaga ggtgcgcccg cgacaacctg ggtcttctgc cccctcgagc ccggggggtc 1380

```

```

gtgcggtggcg ggcaggetcc caggcttggg gcccggtgga agtccatctg acaaccccac 1440
ccaggccagg gtcgaatctg gaatgggagg gctctggcttc agctatcagg gcaccctccc 1500
cagggaattgg aaacggatga cgggcctcta ggcggtcttc tgccacgagc agtttctcat 1560
tactgtctgt ggttaagtcc cctccctcct ttccaaaaat atattacagt cacaccataa 1620
gcacaaacca ggtccagagg tcacctgtga ggagcaaat ccttgtagtc caaatgtat 1680
gagggcgctgg ccacatcagc acttaggaga ggcctctgac aggtccacct cagagccgac 1740
cctccagagc aacttttctg ttgtgaagag gctggttttc tgagtaacgg ttgaaatgtg 1800
caactttcaa cactggaaat ttgcgtgcaa ctagtggagt cagatgctca cactacgaga 1860
agcatgggga acaaaagggt accaaatggg accgtgtaca gtccagtgtt gacaaagggg 1920
tcaatctctt ggcgtaaaag atctcattct ctgccagtg tatagagtgg aatcagccca 1980
ttgtgtaaat atggcctctt tccccagac agctgatctg ctttgtccct ggccagcccc 2040
acttcaggat gagggaggcc ttgctgtcac ccaccacca gtccaactgg gccactttga 2100
cacagcagaa tgccatggga cgagctctct tccgggtcct gtcaatcacc agcaatata 2160
ccttgagaaa tgcattcctc tcccagggcc tcagtttcca catctgtaga atgacggggt 2220
ctatccagaa cctgtctccc tcttacgcac ctctggcagg actttctgga acagggccct 2280
gcaccataag ggggcagaga caaagcccca ggtgtggcat cctggcacc cgtggcactc 2340
agccacttcc cctggtctcc cagatagggc tgtctcaaca caaagcatga acagccagg 2400
ttgaaaaagc caaaaacttt aatttcaaca tgagctacat ggtcatgaac acaatgcaac 2460

```

```

<210> 29
<211> 896
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7474286CB1

```

```

<400> 29
gggaaactga gtccctcacc ccttcaaga ccccgaggcg ctctctgctc cggccctctg 60
agggcctctcg cggcgtctgc ctctccccc ttccgcacc caccggccat aagatgatgt 120
ggtccaactt ctctctgcaa gaggagaacc ggcggcgggg ggccgcgggc cgccggcggt 180
cgacagggga ggcagagtgc ggcgtgacgc cggagcgcca ggggaaggtg aagctgcgc 240
tgctgtctgc cgcgtggggc gccacgtctg cgtgtctgtc cgtgggcaac gagtcttgg 300
tggagctcaa cactacaag gccaacggca gcgcgctgtc cgaagcgccc caactggggc 360
tgtgaaagcg gtgcaccaag cggctgtggc agcgcgagct gcccggtgac agggcaacct 420
cgccgccccc ggagctgcgc ggagaagcaa actgcacct ttttaaattc ttcaccagg 480
gggagaatcg acgcatctt cagagaacca caaagaaga ggtgaatctg gcagctgcgg 540
tgatagcagt ctggggcctg gcagtcatgg ccttggggtg cctctgtatc atcatgtgtc 600
tcagtaaaag tgcagagttc ctgctccagg ttggagcggt ctgcttggc ctctcagacc 660
tgctgtctct ggtgagcctg gaggtgttcc ggcattccgt gaggggcctg ctgcagagag 720
tcagcccgga gccctcccgc gccccacgac tcacctacga gtactcctg tccctggact 780
cgccgctggg ggcggcgctg atcctgctgt tggggggcgg ctgctttctg ctgtctacac 840
tgcttctctg gccctggggg tccctctgtc ccaagcgggg gcaccggggc acctag 896

```

```

<210> 30
<211> 2080
<212> DNA
<213> Homo sapiens

<220>
<221> misc_feature
<223> Incyte ID No: 7472589CB1

```

```

<400> 30
gaagaggagcg cttcccggga ctcggtctgg ctccgaggtc ccgaagccga cgccgccagc 60
tcagcccccgg gggcgggagc aggactgcgc gcacagccc cacttaggag gcgccgctc 120
cgacacgctc attggaagcgc cccggggggt ctctccaccg cttgtctcgg cgtcccgctc 180
ctagggcgccc ggtatccacg gccaccccc ccgcagccc cggcctgttc ttgagagagag 240
tatgagccccc ggggccccgc ggacggccgc caccggcgcg caggggcctt agctcgcgag 300
ccccgcgccc gagagcgggc ggttaaggag cgcgggagcg ggcgagcggt cggggcgcc 360
agagggagcgc cctgtccccc ggcaaccgct gggcgaacgg actcgcgtgt ggcctgagc 420
cgggggagga ggcggagggc cccctctgtc cgggctctgg gaaggcgagc aggggctctg 480

```

```

cgaaggcggc gaggggctcc gcgccgcccc cggacccttg gccaccatcc tcacgtcct 540
gctcccgcgc gggggatgtc gtggcccgcc ccccgagcgc cgcgccggcc ccggggctga 600
gctccggacc atgtctccc gcagccccc gccccgcgc cgcgtagcc gccgcgcct 660
gcgcgcccc tctgtctgtc gctgctgtgc tcgcccgttc cactcaacg aggcaccgg 720
ccgcttcgtg ctgctggcgc cgtctcatcg cctctacctg gtggcggtg ccacgtctt 780
ctcgccgtcc gagagcccc gcgaggcgga gccgcggcg gcgtggggc ccacgtgctg 840
caacttcagc gctgcgcacg gcgtggcgga gccagagctg cgcgccttc tccgcacta 900
cgaggccgcg ctggccgcgc gcgtccgcgc cgacgcgtg cgcgcgcct gggactctcc 960
cgcgcccttc tacttcgtgg gcaccgtgtg gtcaaccata gtgagggaag aaagcccacc 1020
tgtggcgctc accccgggcc gctgtgtctc caacactgac cggtctgtg atctgacctt 1080
taagagttag atcaatattg ccaagaagca ggagcaccga gcaatacagc agactctccc 1140
acgggtttct acagtgtctt cagaagaacc caaggagggt ttccgcatga ccacccccg 1200
gacgttgggc ggaaggcctt tctctatcgc ctacggcgtg ttccgctgac ctgggaacct 1260
cctgttcttc aacctcttcc tggagcgcat catctcgctg ctggccttca tcatgcgcc 1320
ctgcggggag cgcagctgac gccgcagcgc cctgctcccc gccaccttc gcccggtc 1380
cgcgctctgc gagccgaca gctggcggg ctggaagccc tcggtgtacc acgtgctgt 1440
ctgcggggag cgttccgcgc tgctgctgtc ctgctgcgcc tccgcatgt acaccagct 1500
ggaggcgtg gactacgtgc actcgctcta ctctgcttc gtacactta gccacctg 1560
cttcggggac cgtgtgagca gccagcacgc cgcctacgg aaccagggc tctacgcct 1620
ggcacaactc ctctcatcc tgctcgcgct gtgctgcat tactcgctt tcaacgtcat 1680
ctccatcttc atcaagcagg tgctcaactg gatgctgcgc aagctgagct gccgtgctg 1740
cgcgcgctgc tggccgctc ctggccgcgc cctggccgcg cgaatgcca tcacccagg 1800
ctcccggtgc cgcgcgcgc tggccgcgct cggtgccgac ccgcggccc gcgacagca 1860
cgccgagggc cgcgcctctc cgggcgagct catctccatg cgcgacctca cggcctcaa 1920
caaggtgtgc ctggcgctgc tgcagaagca gctgtcggag acggccaacg gctaccgcg 1980
cagcgtgtgc gtcaacacgc gccagaacgc ctctcgggc gccgtgggcg cgctgggcat 2040
catgaacaac cggctggcgc agaccagcgc ctccaggtag 2080

```